HIGH-THROUGHPUT SCREEN FOR
MODULATORS OF CHROMATIN MODIFYING ACTIVITY

TECHNICAL FIELD

5   This disclosure resides in the fields of cellular engineering and drug discovery; in particular, methods and compositions for identifying modulators of chromatin modifying activity.
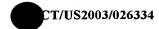
BACKGROUND

The organization of cellular DNA plays a crucial role in the regulation of gene expression.
10   Cellular DNA generally exists in the form of chromatin, a complex comprising nucleic acid and protein. Indeed, most cellular RNAs also exist in the form of nucleoprotein complexes. The nucleoprotein structure of chromatin has been the subject of extensive research, as is known to those of skill in the art. In general, chromosomal DNA is packaged into nucleosomes. A nucleosome comprises a core and a linker. The nucleosome core comprises an octamer of core
15   histones (two each of H2A, H2B, H3 and H4) around which is wrapped approximately 150 base pairs of chromosomal DNA. In addition, a linker DNA segment of approximately 50 base pairs is associated with linker histone H1. Nucleosomes are organized into a higher-order chromatin fiber and chromatin fibers are organized into chromosomes. *See*, for example, Wolffe "Chromatin: Structure and Function" 3rd Ed., Academic Press, San Diego, 1998.

20   Further, cellular chromatin, including nucleosome structure, is organized into a higher order structure of regions or "domains." In those tissues where a given gene or gene cluster is active, the domain is sensitive to DNase I, suggesting that the chromatin of an active domain is in a loose, decondensed configuration that is easily accessible to trans-acting factors (Lawson et al. (1982). *J. Biol. Chem.*, 257:1501-1507; Groudine et al. (1983). *Proc, Natl. Acad. Sci. USA*,
25   80:7551-7555). By contrast, in those tissues where the same gene is not active, the chromatin of the domain is in a tight configuration that is inaccessible to trans-acting factors. Thus, decondensing the higher order chromatin structure of a domain is required before regulatory factors (*e.g.*, transcription factors that bind to specific DNA sequences) can interact with target sequences, thereby determining the transcriptional competence of that domain. Indeed, chromatin structure
30   has been associated with the process of gene regulation *in vivo* and with specific epigenetic states associated with repressed, silenced and induced gene promoters and/or enhancers. (See, *e.g.*, Cheung et al. (2000), *Cell* 103(2):263-271; Strahl & Allis (2000) *Nature* 403(6765):41-45; Jenuwein & Allis (2001) *Science* 293(5532):1074-1080; Wells (2001) *J. Cell Biol.* 154(5):907-907). In particular, it is now thought that these epigenetic states are primarily the result of specific
35   targeted regulatory modifications to the fundamental unit of chromatin structure, the nucleosome.

Researchers have identified various classes of proteins that may be involved in modifying chromatin structure. These chromatin modifying enzymes (CMEs) include histone acetyltransferases (HATs) which modify specific lysine residues in the amino terminal tails of histone proteins. HAT activity has been shown to be exhibited by co-activator proteins, such as

5   p300 and CBP and co-activator complexes such as SAGA exhibit potent HAT activity. (See, Bannister & Kouzarides (1996) *Nature* 384(6610):641-643; Grant et al. (1997) *Genes Dev* 11(13):1640-1650; Kuo et al. (1996) *Nature* 383(6597):269-272; Korzus et al. (1998) *Science* 279(5351):703-707; Li et al. (2000) *Mol. Cell Biol* 20(6):2031-2042; Bhaumik et al. (2001) *Genes Devel.* 15(15):1935-1945; Brown et al. (2001) *Science* 292(5525):2333-2337).

10   Another exempalry class of CMEs includes histone deacteylases, which remove acetyl modifications from histone tails. For example, histone deactelyase activity may play a role in transcriptional co-repression by molecules such as NCoR/SMRT (Laherty et al. (1997) *Cell* 89(3)349-356; Nagy et al. (1997) *Cell* 89(3):373-380; Huang et al. (2000) *Genes Devel.* 14(1):45-54; Li et al. (2000) *EMBO J.* 19(16):4342-4350); Underhill et al. (2000) *J. Biol. Chem.*

15   275(51):40463-40470; Wen et al. (2000) *Proc. Nat'l. Acad. Sci. USA* 97(13):7202-7207; Jones et al. (2001) *J. Biol. Chem.* 276(12):8807-8811); Rb (Brehm (1998) *Nature* 391(6667):597-601; Brehm (1999) *Trends Biochem. Sci.* 24(4):142-145); and Sin3a (Laherty et al. (1997) *Cell* 89(3):349-356; David et al. (1998) *Oncogene* 16(19):2549-2556).

Histone methyltranferases, an additional class of CME, including arginine

20   methyltransferases in the PRMT1 and CARM1 families, target specific arginine residues of H3 and H4 and have been associated with transcriptional activation of nuclear hormone receptors. (See, Ma et al. (2001) *Curr. Biol.* 11(24):1981-1985; Strahl et al. (2001) *Curr. Biol.* 11(12):996-1000; Wang et al. (2001) *Science* 293(5531):853-857; Bauer et al. (2002) *EMBO Rep.* 3(1):39-44). Members of this family contain a SET domain that appears to be integral to their catalytic activity.

25   Six mammalian lysine methyltransferases have been identified, mostly associated with gene repression, which target Lysine 9 of Histone H3 for methylation. This modification appears indicative of areas of compacted chromatin and silenced genes, such as the *S.Pombe* mating type locus (Noma et al. (2001) *Science* 293(5532):1150-1155) or the mammalian inactive female X chromosome (Heard et al. (2001) *Cell* 107(6):727-738). These H3K9 methyltransferases include

30   Suv39H1 (Rea et al. (2000) *Nature* 406(6796):593-599); Suv39H2 (O'Carroll et al. (2000) Mol Cell Biol. Dec; 20(24):9423-33); G9A (Tachibana et al. (2001) *J. Biol. Chem.* 276(27):25309-25317) and SETDB1/ESET (Schultz et al. (2002) *Genes Dev* 16:919-932; Yang et al. (2002) *Oncogene* 21(1):148-152). Set7/9 specifically methylates Lysine 4 of Histone H3 and this modification appears to be associated with transcriptional activation. *See, e.g.,* Nishioka et al.

35   (2002) *Genes Dev* 16:479-489; Wang et al. (2001) *Science* 293(5531):853-857; Zegerman et al. (2002) *J. Biol. Chem.* 277(14):11621-11624.
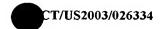
A further exemplary class of CME includes histone kinases, which may phosphorylate serine residues of histones, and corresponding histone phosphatases. Kinases involved in phosphorylation of serine 10 of histone H3 have been shown to be associated with transcriptional activation. Lo *et al.* (2001) *Science* 293:1142-1146.

5 The structure of chromatin can also be altered through the activity of macromolecular assemblies known as chromatin remodeling complexes. *See,* for example, Cairns (1998) *Trends Biochem. Sci.* 23:20-25; Workman et al. (1998) *Ann. Rev. Biochem.* 67:545-579; Kingston et al. (1999) *Genes Devel.* 13:2339-2352 and Murchardt et al. (1999) *J. Mol. Biol.* 293:185-197.

Despite the identification and characterization of certain chromatin modifying enzymes, 10 and progress in elucidating their biological functions, almost no specific CME modulatory (*e.g.,* activator or inhibitor) compounds have been identified. Currently, broad spectrum inhibitors such as sodium butyrate and trichostatin A, in the case of the histone deacetylases, are used to modulate these activities. Such compounds can have genome-wide effects and do not exhibit specificity for a particular CME. Specific compounds have been difficult to identify due to the limitations of the 15 activity assays available, which commonly entail a bulk immunoprecipitation-in vitro activity assay or related assay, precluding the development of any large-scale screening system. Furthermore, due to the essential roles of CMEs in gene regulation, almost all mouse knockouts of CME genes are embryonic lethals, making it difficult to identify genes regulated *in vivo* by the activities of these factors. In sum, there are presently no compositions and methods that allow for fast and 20 efficient screening of molecules that modulate CMEs. Thus, there remains a need for compositions and methods for identifying molecules that modulate CME activity.

SUMMARY

In one aspect, a method to screen for a modulator of a modifying activity is provided. In 25 certain embodiments, the method comprises (a) providing a cell, wherein the cell contains a fusion protein comprising a modifying activity and a DNA-binding activity, wherein the DNA-binding activity is targeted to a reporter gene, and wherein the fusion protein regulates the expression of the reporter gene; (b) contacting the cell with a substance; and (c) assaying expression of the reporter gene, wherein, if expression of the reporter gene is altered in the presence of the substance, 30 compared to the absence of the substance, the substance is a modulator of the modifying activity. The modifying activity can, for example, modify chromosomal DNA or a chromosomal protein (*e.g.,* a histone or a non-histone chromosomal protein). The modifying activity can be, for example, a DNA methyltransferase, a histone acetyl transferase, a histone methyl transferase, a histone deacetylase or a functional fragment of any of the aforementioned enzymes. In any of the 35 methods and compositions described herein, the reporter gene can be, for example, a chromosomal

gene and/or an endogenous gene (*e.g.*, VEGF, H19 and IGF-2). The substance can be, for example, an activator of the modifying activity or an inhibitor of the modifying activity.

In certain embodiments, the DNA-binding activity comprises a naturally-occurring DNA-binding domain. In additional embodiments, one or more engineered zinc fingers provide DNA-

5   binding activity. In any of the methods described herein, the fusion protein can activate expression of the reporter gene or repress expression of the reporter gene. Further, expression can be monitored by assaying mRNA levels encoded by the reporter gene, by assaying expression of protein encoded by the reporter gene and/or by assaying for alteration of expression of the reporter gene that results in a change in phenotype.

10   In any of the methods described herein, the cell can be a plant cell, a mammalian cell, or a human cell. In certain embodiments, the cell comprises a polynucleotide encoding the fusion protein.

In another aspect, disclosed herein is a method to screen for a modulator of a modifying activity. For example, the method comprises (a) providing a cell, wherein the cell contains a fusion

15   protein comprising an interaction domain and a DNA-binding domain, wherein the interaction domain recruits a modifying activity, wherein the DNA-binding domain is targeted to a reporter gene, and wherein the modifying activity regulates the expression of the reporter gene; (b) contacting the cell with a substance; and (c) assaying expression of the reporter gene, wherein, if expression of the reporter gene is altered in the presence of the substance, compared to the absence

20   of the substance, the substance is a modulator of the modifying activity.

In yet another aspect, disclosed herein is a fusion protein comprising a modifying activity (or functional fragment thereof) and a DNA-binding activity, wherein the DNA-binding activity is targeted to a reporter gene, and wherein the modifying activity regulates the expression of the reporter gene. Polynucleotides encoding such fusion proteins are also provided. In certain

25   embodiments, a polynucleotide encoding the fusion protein is the product of an oncogenic chromosomal translocation.

In additional aspects, a fusion protein comprising an interaction domain (or functional fragment thereof) and a DNA-binding domain is provided. In these cases, the interaction domain is capable of recruiting a chromatin modifying activity and the DNA-binding domain is targeted to a

30   reporter gene, such that the recruited modifying activity regulates expression of the reporter gene. Polynucleotides encoding such fusion proteins are also provided. In certain embodiments, a polynucleotide encoding the fusion protein is the product of an oncogenic chromosomal translocation.

For any of the fusion proteins disclosed herein, the modifying activity can comprise, for

35   example, a histone methyltransferase, a histone demethylase, a histone kinase (*e.g.*, SNF1, pp90Rsk, IKK alpha), a histone phosphatase (*e.g.*, PP2A), a histone ubiquitinating enzyme (*e.g.*,

4

RAD6), a histone de-ubiquitinating enzyme, a histone ADP-ribosylating enzyme (*e.g.*, PARP1), a histone de-ribosylating enzyme (*e.g.*, PARG), a histone aminotransferase, a histone deaminase, a histone iminase, a histone de-iminase (*e.g.*, peptidylarginine deiminase, PAD), a DNA aminotransferase, a DNA deaminase, a DNA methyltransferase (*e.g.*, DNMT1, DNMT3a,

5 DNMT3b), a DNA demethylase, a histone acetyl transferase, a histone deacetylase, a histone protease or functional fragments of any of these enzymes. Similarly, for any of the fusion proteins disclosed herein, the interaction domain can recruit a histone methyltransferase, a histone demethylase, a histone kinase, a histone phosphatase, a histone ubiquitinating enzyme, a histone de-ubiquitinating enzyme, a histone ADP-ribosylating enzyme, a histone de-ribosylating enzyme, a

10 histone aminotransferase, a histone deaminase, a histone iminase, a histone de-iminase, a DNA aminotransferase, a DNA deaminase, a DNA methyltransferase, a DNA demethylase, a histone acetyl transferase, a histone deacetylase, a histone protease or a functional fragment of any of these enzymes. *See*, for example, Bannister *et al.* (2002) *Cell* 109:801-806; Lo *et al.* (2001) *Science* 293:1142-1146; Israel (2003) *Nature* 423:596-597; Nowak *et al.* (2003) *Mol. Cell. Biol.* 23:6129-

15 6138; Sun *et al.* (2002) *Nature* 418:104-108; Brummelkamp *et al.* (2003) *Nature* 424:797-801; Kraus *et al.* (2003) *Cell* 113:677-683; Nakashima *et al.* (2002) *J. Biol. Chem.* 277:49,562-49,568; Neuberger *et al.* (2003) *Trends Biochem. Sci.* 28:305-312; Rountree *et al.* (2001) *Oncogene* 20:3156-3165; and Dino-Rockel *et al.* (2002) *J. Struct. Biol.* 140:189-199.

To the extent that the modifying activities mentioned in the previous paragraph result in

20 covalent histone modification, chromatin remodeling and/or modulation of gene expression, the disclosed fusion proteins can also be used for these purposes. For example, the enzyme peptidylarginine deiminase (PAD) converts arginine to citrulline in proteins. Four types of PADs, PAD I, PAD II, PAD III and PAD V, exist in humans. PAD V converts arginine to citrulline in histones H1, H2, H3 and H4, and this modification can antagonize the transcription-stimulatory

25 activities of nuclear hormone receptors. Thus, for a particular gene whose expression (*e.g.*, transcription) is activated by a nuclear hormone receptor, a fusion protein comprising a PAD (e.g., PAD V), or functional fragment thereof, and a zinc finger binding domain targeted to the gene, can be used to repress and/or block nuclear hormone receptor-stimulated transcription of the targeted gene.

30 Methylation of lysine 9 on histones H3 and H4 can result in repression of gene expression, if the methylated histones are present in a nucleosome in the vicinity of the gene. Accordingly, fusions of a ZFP binding domain (targeted to a gene of interest) and the catalytic domain of a histone methyltransferase (*e.g.*, Suv39H1, G9A), when expressed in a cell, can be used for targeted repression of gene expression.

35 Polynucleotides encoding the fusion proteins disclosed herein, as well as cells comprising said polynucleotides and fusion proteins, are also provided.

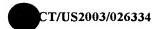Also disclosed herein is a method of screening for a compound that modulates the activity of a chromatin modifying enzyme, the method comprising the steps of: (a) contacting a cell with the compound, wherein the cell comprises a fusion protein comprising a zinc finger protein (ZFP) and a functional chromatin modifying enzyme or fragment thereof and wherein the ZFP binds to a

5   reporter gene; and (b) determining the level of expression of the reporter gene. The fusion protein can be provided as a polypeptide and/or as a polynucleotide encoding the fusion protein. The cell can be stably or transiently transfected with the polynucleotide encoding the fusion protein. In yet other embodiments, the polynucleotide further comprises an inducible promoter (*e.g.,* a tetracycline-inducible promoter) operably linked to the polynucleotide encoding the fusion protein.

10   In any of the methods described herein, exemplary chromatin modifying enzymes can be selected from the group consisting of a histone methyltransferase (HMT) (*e.g.,* lysine or arginine HMTases such as those which methylate H3 lysine 4 (H3K4), H3 lysine 9 (H3K9), H3 lysine 27 (H3K27), H3 lysine 36 (H3K36) and H4 lysine 20 (H4K20), a histone deacetylase (HDAC) (*e.g.,* Sir2 or any of HDACs 1-11), a DNA methyltransferase (DNMT) (*e.g.,* DNMT1, DNMT3a,

15   DNMT3b), a histone acetyltransferase (HAT) (*e.g.,* p300, CBP, pCAF) and functional fragments thereof. Similarly, exemplary interaction domains include, but are not limited to, v-*erbA* and protein component of the NCoR, Sin3A, or Rb complexes, as well as functional fragments thereof.

In a still further aspect, polynucleotides encoding any of the fusion proteins disclosed herein are provided.

20   In a further aspect, cells comprising any of the fusion proteins or any polynucleotides described herein are disclosed.

These and other embodiments will be readily apparent to the skilled artisan in view of the disclosure herein.

25   BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A shows schematic representations of the histone methyltransferases G9A and SUV39H1, indicating the regions of each protein used as a ZFP fusions in relation to known structural features.

Figure 1B shows that ZFP-HMT fusions repress VEGF-A protein expression. HEK293

30   cells transfected with the indicated plasmids were assayed for VEGF-A protein production using a human VEGF-A ELISA assay kit as described in Example 5.

Figure 1C shows that ZFP-HMT fusions repress VEGF-A mRNA levels. *VEGF-A* mRNA levels were determined by real time PCR (TaqMan) after expression of the indicated ZFPs for 72hrs by transient transfection. The *VEGF-A* mRNA levels were normalized relative to an internal

35   control of *GAPDH* mRNA, and are expressed as this ratio.

Figure 1D shows that ZFP-HMT fusions do not repress a VEGF-reporter plasmid. HEK293 cells transfected with the VEGF-A promoter firefly luciferase reporter plasmid (800ng) together with the indicated ZFP expression plasmid (100ng) and TK renilla luciferase control reporter (50ng), were assayed for luciferase activity using the Dual-Luciferase assay 72hrs post
5   transfection.

Figure 1E shows a Western blot of ZFP-HMT constructs. HEK293 cells were transfected with the indicated plasmids and whole cell lysates prepared 72hrs post transfection. Extracts were resolved by SDS-PAGE and immunoblotted using an anti HA-epitope tag antibody. Equal protein levels were loaded in each lane. A band of roughly equivalent mobility to ZFP-Suv Del 76 cross-
10   reacts in HEK293 extracts (see also Fig. 2A Panel II).

Figure 2A. Panel I shows that ZFP-HMTs fusions are catalytically active *in vitro*. Extracts from HEK293 cells transfected with the indicated plasmids were immunoprecipitated with either an anti-HA epitope tag antibody (Anti-HA IP) or a IgG control antibody (IgG control IP), and assayed for histone methyltransferase activity. Panel II shows a Western blot of extracts used in
15   the HMT assay (Panel I). See also Fig. 1E.

Figure 2B provides schematics showing the locations of the amino acid substitution mutants generated within the HMT catalytic core motif of SUV39H1 (Suv Del 76 wild type: SEQ ID NO:1; Suv Del 76 mutant A: SEQ ID NO:2; Suv Del 76 mutant B: SEQ ID NO:3; Suv Del 76 mutant AB: SEQ ID NO:4).

20   Figure 2C shows that ZFP-HMT fusions are dependent upon their catalytic HMT activity for repression function *in vivo*. HEK293 cells transfected with the plasmids indicated were assayed for VEGF-A mRNA by quantitative RT-PCR (TaqMan) as described in Fig. 1C.

Figure 2D shows that expression levels of the ZFP-HMT mutants are comparable to that of the wild type fusion protein. Extracts from HEK293 cells transfected with the indicated plasmids
25   were immunoblotted as in Fig. 1E.

Figure 3A shows a schematic of the VEGF-A gene promoter indicating the ZFP binding sites relative to the transcriptional start site. Positions of the CHIP primer-pairs used in Figure 4 are also indicated (gray boxes).

Figure 3B shows that HMT and V-ErbA functional domains repress VEGF-A transcription
30   when linked to either ZFP-A or ZFP-B. The indicated combinations of ZFP-A, ZFP-B and functional domain were assayed for VEGF-A mRNA levels by quantitative RT-PCR (TaqMan) as described in Fig. 1C.

Figure 3C shows that simultaneous targeting of both an HMT and v-ErbA enhances repression of VEGF-A transcription. HEK293 cells transfected with the plasmids indicated were
35   assayed for VEGF-A mRNA by quantitative RT-PCR (TaqMan) as described in Fig. 1C.

Figure 3D shows that dual targeting of the same functional domain with both ZFP-A and ZFP-B does not enhance VEGF-A repression. HEK293 cells transfected with the plasmids indicated were assayed for VEGF-A mRNA by quantitative RT-PCR (TaqMan) as described in Fig. 1C.

5   Figure 4A shows that ZFP-HMT fusions methylate H3K9 promoter nucleosomes at the VEGF-A locus and require HMT catalytic activity to methylate H3K9 at the VEGF-A promoter. HEK293 cells transfected with the indicated plasmids were assayed for H3K9 methylation by ChIP with primers specific for the ZFP proximal +400 region. Enrichment was quantified by RT-PCR. Results are expressed as the fold-increase of the ratio to the GAPDH control relative to the results

10   for non-transfected cells, the value of which is arbitrarily set to 1. The same samples were analyzed with primers specific the p16 locus as a second internal control (light gray bars). No enrichment was observed with pre-immune serum.

Figure 4B shows that ZFP-G9A induces the spread of H3K9 methylation across the VEGF-A promoter in vivo. HEK293 cells transfected with the plasmids indicated were assayed for

15   methylation of H3K9 by ChIP with primers specific for the regions centered on +400, +1, and – 500. Samples were treated as Fig.4A.

Figure 4C shows that ZFP-SUV Del 76 is dependent upon its catalytic HMT activity for spreading of H3K9 methylation across VEGF-A promoter in vivo. HEK293 cells transfected with the plasmids indicated were assayed for H3K9 by ChIP with primers specific for the regions

20   centered on +400, +1, and –500. Samples were treated as in Fig.4A.

Figure 5 is a graph depicting levels of secreted VEGF protein in foci of stably transformed cells inducibly expressing a ZFP-CME fusion. Isolated clonal populations of TREx U2OS cells which had been stably transformed with a TREx regulated ZFP-G9A expression vector were screened for Dox-dependent VEGF-A repression, as assayed by ELISA.

25   Figure 6 is a graph depicting levels of secreted VEGF protein in foci of stably transformed cells inducibly expressing a ZFP-CME fusion in foci of TREx U2OS cells. Isolated clonal populations of TREx U2OS cells which had been stably transformed with a TREx regulated ZFP-Suv del 76 expression vector were screened for Dox-dependent VEGF-A repression, as assayed by ELISA.

30   Figure 7 is a graph depicting repression of IGF2 transcription by VOP32-G9A in TREx-inducible U2OS cells.

Figure 8 is a graph depicting repression of H19 transcription by VOP32-G9A in TREx-inducible U2OS cells.

Figure 9 is a graph depicting repression of IGF2 transcription by VOP32-Suvdel76 in

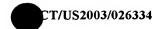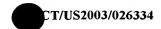35   TREx-inducible U2OS cells.

Figure 10 is a graph depicting repression of H19 transcription by VOP32-Suvdel76 in TREx-inducible U2OS cells.

Figure 11 is a graph depicting levels of VEGF mRNA, normalized to levels of GAPDH mRNA, in human HEK 293 cultured cells transfected with various plasmids, as indicated. Lane 1:

5   Cells were transfected with 1.5ug of an expression plasmid encoding the Vop32 zinc finger DNA binding protein fused with full length PAD V enzyme at its carboxy terminus. Lane 2: Cells were transfected with 1.5ug of an expression plasmid encoding the Vop30 zinc finger DNA binding protein fused with full length PAD V enzyme at it's carboxy terminus. Lane 3: Cells were transfected with 1.5ug of an expression plasmid encoding the 5499 zinc finger DNA binding

10   protein fused with full length PAD V enzyme at its carboxy terminus. 5499 does not target or regulate the VEGF-A gene and acts as a control for the function of Vop30 PAD V. Lane 4: Cells were transfected with 1.25ug of an expression plasmid control pCDNA4.1, which functions as an empty expression vector control, together with 250ng of an expression vector encoding the Vop32 zinc finger DNA binding protein fused with the ER alpha LBD at its carboxy terminus. Cells were

15   stimulated with the ER alpha ligand beta-estradiol (indicated as "B est" in the figure) 24hours post-transfection. Lane 5: Cells were transfected with 1.25ug of an expression plasmid encoding the Vop32 zinc finger DNA binding protein fused with full length PAD V enzyme at its carboxy terminus, together with 250ng of an expression vector encoding the Vop32 zinc finger DNA binding protein fused with the ER alpha LBD at its carboxy terminus. Cells were stimulated with

20   the ER alpha ligand 24hours post transfection. Lane 6: Cells were transfected with 1.25ug of an expression plasmid encoding the 5499 zinc finger DNA binding protein fused with full length PAD V enzyme at its carboxy terminus, together with 250ng of an expression vector encoding the Vop32 zinc finger DNA binding protein fused with the ER alpha LBD at its carboxy terminus. 5499 does not target or regulate the VEGF-A gene and acts as a control for the function of Vop30

25   PAD V. Cells were stimulated with the ER alpha ligand 24hours post transfection. Lane 7: Cells were transfected with 1.5ug of an expression plasmid control pCDNA4.1, which functions as an expression vector control. Lane 8: Cells were transfected with 1.5ug of an expression plasmid control encoding the Green Fluorescent protein (GFP) fused at its carboxy terminus to the PAD V enzyme, which functions as a PAD V expression control. Lane 9: Cells were mock transfected

30   with Lipofectamine 2000 reagent in the absence of DNA and stimulated 24 hours post transfection with the ER alpha ligand Beta Estradiol.

DETAILED DESCRIPTION

Overview

35   The compositions and methods disclosed herein include novel assays (e.g., cell-based assays) for screening candidate compounds for their ability to modulate chromatin-modifying

9

enzymes (CMEs). Identification of these CME modulators is useful in a variety of instances, for example, in diagnosing and/or treating a variety of conditions or disease states. However, since currently available assay systems are based on immunoprecipitation and/or *in vitro* enzymatic activity assays, they do not readily allow for identification of such modulators. The currently

5   available assays are problematic for one or more of the following reasons: they generally involve the use of potentially hazardous radioactive substrates (*e.g.*, tritium or ^{14}C); they are not scale-able and are difficult to utilize and standardize; and/or they give no indication of *in vivo* function or cellular toxicity. Thus, known experimental approaches are not very amenable to high throughput screening for CMEs.

10   Part of the reason that it is difficult to screen for modulators of chromatin modifying activity is that most of the enzymes which catalyze covalent chromatin modifications do not bind directly to DNA, but are targeted to specific genomic regions through recruitment by other DNA-binding proteins. Due to the complex regulation of the formation and dissociation of such complexes *in vivo*, it has been difficult to identify specific modulators of chromatin modifying

15   activities. The present disclosure addresses these problems by providing fusions of chromatin modifying activities to targeted DNA-binding domains, allowing a particular chromatin modifying activity to be directed to a specific reporter gene and thereby regulate the expression of that gene. This then provides an assay for compounds which modulate the targeted chromatin modifying activity, as such compounds will elicit changes in the regulation of expression of the reporter gene

20   by the targeted chromatin modifying activity.

It should also be pointed out that, since chromatin modifying activities normally act on nucleosomes in cellular chromatin, plasmid-based reporters (which are not assembled into normal chromatin when introduced into cells) are not regulated by chromatin modifying enzymes. Consequently, traditional transient reporter-based assays are not capable of detecting chromatin

25   modifying activities or their modulation.

In one embodiment, cells are transfected with a polynucleotide encoding a chimeric protein. The chimeric protein comprises a fusion of a full-length or a fragment of a CME (*e.g.*, the catalytic domain) and a DNA binding domain (*e.g.* a naturally-occurring DNA-binding domain or an engineered ZFP DNA-binding domain), which targets one or more endogenous gene(s) that can

30   act as reporter(s). Expression of a gene encoding a CME-ZFP fusion is optionally under the control of an inducible promoter. In addition, the chimeric (fusion) proteins are capable of significantly up- or down-regulating the transcription of these endogenous 'reporter' genes, which in turn allows for the direct and rapid screening of molecules that affect the CME (*e.g.*, by modulating its activity). The cells can be stably or transiently transfected with a ZFP-CME-

35   encoding construct.

The CME-ZFP fusion proteins or polynucleotides encoding these proteins are introduced into cells (*e.g.*, via stable or transient transfection). The effect of a candidate compound on the molecular target (*e.g.*, the CME) can be determined by comparing expression of the reporter gene(s) in the fusion protein-containing cells in the presence and absence of the compound. To

5   control for non-specific effects of a compound, fusion protein-containing cells exposed to a compound can be compared to cells which do not contain the fusion protein but are exposed to the same compound. In one embodiment, the aforementioned two populations of cells can be provided by a single cell line containing a polynucleotide encoding the fusion protein under the control of an inducible promoter.

10   The practice of conventional techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA, bioinformatics, genomics and related fields are well-known to those of skill in the art and are discussed, for example, in the following literature references: Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, Third edition, Cold Spring Harbor Laboratory Press, 2001; Ausubel et al., CURRENT PROTOCOLS

15   IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.)

20   Humana Press, Totowa, 1999, all of which are incorporated by reference in their entireties.

The disclosures of all patents, patent applications and publications mentioned herein are hereby incorporated by reference in their entireties.

Definitions

25   The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties. In general, an analogue of a

30   particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T. Thus, the term polynucleotide sequence is the alphabetical representation of a polynucleotide molecule. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

Chromatin is the nucleoprotein structure comprising the cellular genome. "Cellular

35   chromatin" comprises nucleic acid, primarily DNA, and protein, including histones and non-histone chromosomal proteins. The majority of eukaryotic cellular chromatin exists in the form of

nucleosomes, wherein a nucleosome core comprises approximately 150 base pairs of DNA associated with an octamer comprising two each of histones H2A, H2B, H3 and H4; and linker DNA (of variable length depending on the organism) extends between nucleosome cores. A molecule of histone H1 is generally associated with the linker DNA. For the purposes of the present disclosure, the term "chromatin" is meant to encompass all types of cellular nucleoprotein, both prokaryotic and eukaryotic. Cellular chromatin includes both chromosomal and episomal chromatin.

A "chromosome" is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the cell. The genome of a cell can comprise one or more chromosomes.

An "episome" is a replicating nucleic acid, nucleoprotein complex or other structure comprising a nucleic acid that is not part of the chromosomal karyotype of a cell. Examples of episomes include plasmids and certain viral genomes.

Typical "control elements" include, but are not limited to, transcription promoters, transcription enhancer elements, silencers, locus control regions, insulators, boundary elements, matrix attachment regions, replication origins, cis-acting transcription regulating elements (transcription regulators, e.g., a cis-acting element that affects the transcription of a gene, for example, a region of a promoter with which a transcription factor interacts to modulate expression of a gene), transcription termination signals, as well as polyadenylation sequences (located 3' to the translation stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), translation enhancing sequences, and translation termination sequences. Transcription promoters can include inducible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, small molecule, drug, etc.), repressible promoters (where expression of a polynucleotide sequence operably linked to the promoter is repressed by an analyte, cofactor, regulatory protein, small molecule, drug, etc.), and constitutive promoters, which are characterized by a constant level of activity in the absence of inducing or repressing substances.
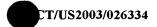
Techniques for determining nucleic acid and amino acid "sequence identity" also are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion. In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied
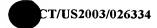
12

by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical

5   Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, WI) in the "BestFit" utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI).

10   A preferred method of establishing percent identity in the context of the present disclosure is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of

15   six). From the data generated the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by =

20   HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address: http://www.ncbi.nlm.gov/cgi-bin/BLAST. When claiming sequences relative to sequences described herein, the range of desired degrees of sequence identity is approximately 80% to 100% and any integer value therebetween. Typically the percent identities between the disclosed

25   sequences and the claimed sequences are at least 70-75%, preferably 80-82%, more preferably 85-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence identity to the reference sequence (i.e., the sequences disclosed herein).

Alternatively, the degree of sequence similarity between polynucleotides can be determined by hybridization of polynucleotides under conditions that allow formation of stable duplexes between

30   homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 70%-75%, preferably 80%-82%, more preferably 85%-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence identity to each other, or to a reference sequence, over a defined length of the molecules,

35   as determined using the methods above. As used herein, substantially homologous also refers to sequences showing complete identity to a specified DNA or polypeptide sequence. DNA sequences

that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; DNA Cloning: A Practical Approach, editor, D.M. Glover (1985) Oxford; Washington, DC; IRL Press; Nucleic Acid

5   Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins (1985) Oxford; Washington, DC; IRL Press.

"Selective hybridization" of two nucleic acid fragments can be determined as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A nucleic acid sequence that is partially identical to a

10   target molecule will at least partially inhibit the hybridization of a completely identical sequence to the target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of

15   selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

20   When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a duplex or "hybrid" molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" hybridization conditions typically hybridizes under conditions

25   that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/target

30   hybridization, where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

Conditions for hybridization are well known to those of skill in the art. Hybridization stringency refers to the degree to which hybridization conditions disfavor the formation of duplexes

35   containing mismatched nucleotides, with higher stringency correlated with a lower tolerance for mismatches. Factors that affect the stringency of hybridization are well-known to those of skill in the
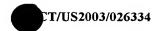
art and include, but are not limited to, temperature, pH, ionic strength, and concentration of organic solvents such as, for example, formamide and dimethylsulfoxide. As is known to those of skill in the art, hybridization stringency is increased by higher temperatures, lower ionic strength and lower solvent concentrations.

5   With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., dextran sulfate, and polyethylene glycol),

10   hybridization reaction temperature and time parameters, as well as varying wash conditions. The selection of a particular set of hybridization conditions is conducted following standard methods in the art (see, for example, Sambrook, *et al., supra*).

The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino

15   acids are chemical analogues or modified derivatives of corresponding naturally-occurring amino acids.

A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule' (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, *etc.*) and/or it can bind

20   to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

A "zinc finger DNA binding protein" is a protein or segment within a larger protein that binds DNA in a sequence-specific manner as a result of stabilization of protein structure through coordination

25   of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

A "designed" zinc finger protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of

30   existing ZFP designs and binding data. A "selected" zinc finger protein is a protein not found in nature whose production results primarily from an empirical process such as phage display. See *e.g.*, US 5,789,538; US 6,007,988; US 6,013,453; US 6,140,081; US 6,140,466; WO 95/19431; WO 96/06166 and WO 98/54311. Both designed and selected ZFPs are examples of "engineered" ZFPs.

35   The term "naturally-occurring" is used to describe an object that can be found in nature, as distinct from being artificially produced by humans. Examples include naturally-occurring zinc fingers

(*e.g.*, a zinc finger that is encoded by the genome of an organism, as opposed to having been designed or selected), and naturally-occurring zinc finger proteins (*e.g.*, a protein comprising multiple zinc fingers wherein the sequence of the entire protein, including the sequence and location of the zinc fingers in the protein, is encoded by the genome of an organism). For the purposes of the present

5   disclosure, a protein comprising a collection of naturally-occurring zinc fingers, which are not normally present together in a naturally-occurring ZFP and/or which are not present in the order in which they occur in a naturally-occurring ZFP, is not a naturally-occurring protein, but is considered to be a type of engineered ZFP.

Nucleic acid or amino acid sequences are "operably linked" (or "operatively linked") when

10   placed into a functional relationship with one another. For instance, a promoter or enhancer is operably linked to a coding sequence if it regulates, or contributes to the modulation of, the transcription of the coding sequence. Operably linked DNA sequences are typically joined in *cis* and can be contiguous, and operably linked amino acid sequences are typically contiguous and in the same reading frame. However, since enhancers generally function when separated from the promoter by up to several

15   kilobases or more and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous. Similarly, certain amino acid sequences that are non-contiguous in a primary polypeptide sequence may nonetheless be operably linked due to, for example folding of a polypeptide chain.

With respect to fusion polypeptides, the term "operatively linked" can refer to the fact that each

20   of the components performs the same function in linkage to the other component as it would if it were not so linked. For example, with respect to a fusion polypeptide in which a ZFP DNA-binding domain is fused to a transcriptional activation domain (or functional fragment thereof), the ZFP DNA-binding domain and the transcriptional activation domain (or functional fragment thereof) are in operative linkage if, in the fusion polypeptide, the ZFP DNA-binding domain portion is able to bind its target site

25   and/or its binding site, while the transcriptional activation domain (or functional fragment thereof) is able to activate transcription.

A "functional fragment" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid whose sequence is not identical to the full-length protein, polypeptide or nucleic acid, yet retains the same function as the full-length protein, polypeptide or nucleic acid. A

30   functional fragment can possess more, fewer, or the same number of residues as the corresponding native molecule, and/or can contain one or more amino acid or nucleotide substitutions. Methods for determining the function of a nucleic acid (*e.g.*, coding function, ability to hybridize to another nucleic acid, binding to a regulatory molecule) are well known in the art. Similarly, methods for determining protein function are well known. For example, the DNA-binding function of a

35   polypeptide can be determined, for example, by filter-binding, electrophoretic mobility-shift, or immunoprecipitation assays. See Ausubel *et al., supra*. The ability of a protein to interact with
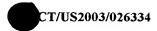
another protein can be determined, for example, by co-immunoprecipitation, two-hybrid assays or complementation, both genetic and biochemical. See, for example, Fields *et al.* (1989) *Nature* 340:245-246; U.S. Patent No. 5,585,245 and PCT WO 98/44350.

5   "Specific binding" between, for example, a ZFP and a specific target site means a binding affinity (*i.e*, K_d) of at least $1 \times 10^6 \text{ M}^{-1}$.

A "fusion molecule" is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same chemical type of molecule, or can be different chemical types of molecules. Examples of the first type of fusion molecule include, but are not limited to, fusion polypeptides (for example, a fusion between a ZFP DNA-binding domain and a

10   protein that exhibits chromatin modifying activity) and fusion nucleic acids (for example, a nucleic acid encoding a ZFP-CME fusion polypeptide). Examples of the second type of fusion molecule include, but are not limited to, a fusion between a triplex-forming nucleic acid and a polypeptide, and a fusion between a minor groove binder and a nucleic acid.

An "exogenous molecule" is a molecule that is not normally present in a cell, but can be

15   introduced into a cell by one or more genetic, biochemical or other methods. Normal presence in the cell is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell.

20   An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally functioning endogenous molecule.

An exogenous molecule can be, among other things, a small molecule, such as is generated by a combinatorial chemistry process, or a macromolecule such as a protein, nucleic acid, carbohydrate, lipid, glycoprotein, lipoprotien, polysaccharide, any modified derivative of the above

25   molecules, or any complex comprising one or more of the above molecules. Nucleic acids include DNA and RNA, can be single- or double-stranded; can be linear, branched or circular; and can be of any length. Nucleic acids include those capable of forming duplexes, as well as triplex-forming nucleic acids. See, for example, U.S. Patent Nos. 5,176,996 and 5,422,251. Proteins include, but are not limited to, DNA-binding proteins, transcription factors, chromatin remodeling factors,

30   methylated DNA binding proteins, polymerases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, integrases, recombinases, ligases, topoisomerases, gyrases and helicases.

An exogenous molecule can be the same type of molecule as an endogenous molecule, *e.g.*, protein or nucleic acid (*e.g.*, an exogenous gene). For example, an exogenous nucleic acid can

35   comprise an infecting viral genome, a plasmid or episome introduced into a cell, or a chromosome that is not normally present in the cell. Methods for the introduction of exogenous molecules into
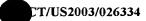
17

cells are known to those of skill in the art and include, but are not limited to, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment, calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer.

5   By contrast, an "endogenous molecule" is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally occurring episomal nucleic acid. Additional endogenous molecules can include endogenous genes and endogenous proteins, for example, transcription factors and components of

10   chromatin remodeling complexes.

       A "gene," for the purposes of the present disclosure, includes a DNA region encoding a gene product (see below), as well as all DNA regions that regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators,

15   translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

       An "endogenous gene" is a gene that is native to a cell, which is in its normal genomic and chromatin context and which is not heterologous to the cell. Endogenous genes can be cellular,

20   microbial or viral. Endogenous microbial and viral genes refer to genes that are part of a naturally-occurring microbial or viral genome in a microbially- or virally-infected cell. The microbial or viral genome can be extrachromosomal, or it can be integrated into the host chromosome(s).

       "Gene expression" refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (*e.g.*, mRNA, tRNA, rRNA,

25   antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of an mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

       "Gene activation" and "augmentation of gene expression" refer to any process which results in

30   an increase in production of a gene product. A gene product can be either RNA (including, but not limited to, mRNA, rRNA, tRNA, enzymatic RNA and structural RNA) or protein. Accordingly, gene activation includes those processes that increase transcription of a gene and/or translation of a mRNA. Examples of gene activation processes which increase transcription include, but are not limited to, those which facilitate formation of a transcription initiation complex, those which increase transcription

35   initiation rate, those which increase transcription elongation rate, those which increase processivity of transcription and those which relieve transcriptional repression (by, for example, blocking the binding

of a transcriptional repressor). Gene activation can constitute, for example, inhibition of repression as well as stimulation of expression above an existing level. Examples of gene activation processes that increase translation include those that increase translational initiation, those that increase translational elongation and those that increase mRNA stability. In general, gene activation comprises any detectable

5   increase in the production of a gene product, preferably an increase in production of a gene product by about 2-fold, more preferably from about 2- to about 5-fold or any integral value therebetween, more preferably between about 5- and about 10-fold or any integral value therebetween, more preferably between about 10- and about 20-fold or any integral value therebetween, still more preferably between about 20- and about 50-fold or any integral value therebetween, more preferably between about 50- and

10   about 100-fold or any integral value therebetween, more preferably 100-fold or more.

"Gene repression" and "inhibition of gene expression" refer to any process that results in a decrease in production of a gene product. A gene product can be either RNA (including, but not limited to, mRNA, rRNA, tRNA, enzymatic RNA and structural RNA) or protein. Accordingly, gene repression includes those processes that decrease transcription of a gene and/or translation of a mRNA.

15   Examples of gene repression processes which decrease transcription include, but are not limited to, those which inhibit formation of a transcription initiation complex, those which decrease transcription initiation rate, those which decrease transcription elongation rate, those which decrease processivity of transcription and those which antagonize transcriptional activation (by, for example, blocking the binding of a transcriptional activator). Gene repression can constitute, for example, prevention of

20   activation as well as inhibition of expression below an existing level. Examples of gene repression processes that decrease translation include those that decrease translational initiation, those that decrease translational elongation and those that decrease mRNA stability. Transcriptional repression includes both reversible and irreversible inactivation of gene transcription. In general, gene repression comprises any detectable decrease in the production of a gene product, preferably a decrease in

25   production of a gene product by about 2-fold, more preferably from about 2- to about 5-fold or any integral value therebetween, more preferably between about 5- and about 10-fold or any integral value therebetween, more preferably between about 10- and about 20-fold or any integral value therebetween, still more preferably between about 20- and about 50-fold or any integral value therebetween, more preferably between about 50- and about 100-fold or any integral value therebetween, more preferably

30   100-fold or more.

"Modulation" of gene expression includes both gene activation and gene repression. Modulation can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, e.g., changes in RNA or protein levels; changes in protein activity; changes in product levels; changes in downstream gene expression; changes in

35   transcription or activity of reporter genes such as, for example, luciferase, CAT, beta-galactosidase, or GFP (see, e.g., Mistili & Spector, (1997) Nature Biotechnology 15:961-964); changes in signal
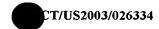
19

transduction; changes in phosphorylation and dephosphorylation; changes in receptor-ligand interactions; changes in concentrations of second messengers such as, for example, cGMP, cAMP, IP$_3$, and Ca2$^+$; changes in cell growth, changes in neovascularization, and/or changes in any functional effect of gene expression. Measurements can be made *in vitro, in vivo*, and/or *ex vivo*. Such functional effects

5   can be measured by conventional methods, *e.g.*, measurement of RNA or protein levels, measurement of RNA stability, and/or identification of downstream or reporter gene expression. Readout can be by way of, for example, chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays; changes in intracellular second messengers such as cGMP and inositol triphosphate (IP$_3$); changes in intracellular calcium levels; cytokine release, and the like.

10   "Eucaryotic cells" include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells.

A "regulatory domain" or "functional domain" refers to a protein or a polypeptide sequence that has transcriptional modulation activity. In one embodiment, a regulatory domain is covalently or non-covalently linked to a ZFP to modulate transcription of a gene of interest. Alternatively, a ZFP can act

15   alone, without a regulatory domain, to modulate transcription. Furthermore, transcription of a gene of interest can be modulated by a ZFP linked to multiple regulatory domains. In addition, a regulatory domain can be linked to any DNA-binding domain having the appropriate specificity to modulate the expression of a gene of interest.

A "target site" or "target sequence" is a sequence that is bound by a binding protein or binding

20   domain such as, for example, a ZFP. Target sequences can be nucleotide sequences (either DNA or RNA) or amino acid sequences. By way of example, a DNA target sequence for a three-finger ZFP is generally either 9 or 10 nucleotides in length, depending upon the presence and/or nature of cross-strand interactions between the ZFP and the target sequence.

The term "recombinant," when used with reference to a cell, indicates that the cell

25   replicates an exogenous nucleic acid, or expresses a peptide or protein encoded by an exogenous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also contain genes found in the native form of the cell wherein the genes are modified and re-introduced into the cell. A recombinant cell can · comprise an unmodified cellular gene which has been introduced into the cell for the purpose, *e.g.*,

30   of overexpression. Expression of such an unmodified gene may be under the control of its normal cellular regulatory sequences or heterologous regulatory sequences. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques.

35   A "recombinant expression cassette," "expression cassette" or "expression construct" is a nucleic acid construct, generated recombinantly or synthetically, that has control elements that are
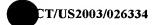
capable of effecting expression of a structural gene that is operatively linked to the control

elements in hosts compatible with such sequences. Expression cassettes include at least promoters

and optionally, transcription termination signals. Typically, the recombinant expression cassette

includes at least a nucleic acid to be transcribed (*e.g.*, a nucleic acid encoding a desired

5   polypeptide) and a promoter. Additional factors necessary or helpful in effecting expression can

also be used as described herein. For example, an expression cassette can also include nucleotide

sequences that encode a signal sequence that directs secretion of an expressed protein from the host

cell, nuclear localization signals and/or epitope tags. Transcription termination signals, enhancers,

and other nucleic acid sequences that influence gene expression, can also be included in an

10   expression cassette.

"K_d" refers to the dissociation constant for the compound, i.e., the concentration of a

compound (e.g., a zinc finger protein) that gives half maximal binding of the compound to its

target (i.e., half of the compound molecules are bound to the target) under given conditions (i.e.,

when [target] $\ll K_d$), as measured using a given assay system (see, e.g., U.S. Patent No.

15   5,789,538). The assay system used to measure the K_d should be chosen so that it gives the most

accurate measure of the actual K_d of the ZFP. Any assay system can be used, as long is it gives an

accurate measurement of the actual K_d of the ZFP.

Chromatin Modifying Molecules

20   The present disclosure provides compositions and methods that allow for the identification

of modulators (*e.g.*, inhibitors or activators) of any type of chromatin modifying activity, acting on

either chromosomal DNA and/or chromosomal proteins, including, for example, chromatin

modifying enzymes such as histone acetyl transferases, histone deacetylases, histone

methyltransferases, DNA methyltransferases, as well as proteins involved in phosphorylation,

25   ubiquitination, glycosylation and ADP-ribosylation of chromatin.

A. Histone methyltransferases (HMTs)

At least two different families of HMTs have been identified thus far, including lysine

methyltransferases and arginine methyltransferases. Lysine methyltransferases methylate specific

30   lysine residues within histone H3 and H4. Methylation of H3 lysine 4 appears to code for gene

activation, in certain cases, and silencing at telomeres (chromosome ends) in others. Most other

lysine methylation tags – H3K9, H3K27 and H3K36 appear to be repressive signals associated

with gene repression and transcriptional silencing. Most of the proteins of the HMT family contain

the SET protein domain, in which catalytic (*i.e.*, methyltransferase) activity appears to reside.

35   Recent studies link histone methylation with long-term epigenetic repression, achieved in part by

subsequent DNA methylation in chromatin comprising methylated histones. Aberrant versions or

translocations of HMT proteins, that are linked to cancer, are also being identified. Additionally, HMT activities are utilized and required for the function of known tumor suppressor genes. This link has provoked significant interest in specific inhibitors of these activities as anti-cancer agents.

5   Approximately 5 arginine methyltransferases have been identified so far (*e.g.*, PRMT1, CARM1) and all of these methylate specific arginine residues of the histone tails of histone H3 or H4. In addition, all are associated with transcriptional coactivators and with gene activation *in vivo*.

B. Histone Deacetylases (HDACs)

10   Histone deacteylases (HDACs) catalyze the removal of acetyl groups from modified lysine residues (that have been acetylated by HATs) within histones, which apparently lowers DNA accessibility and promotes higher order chromatin compaction. Thus, these proteins are associated with transcriptional repression and gene silencing, with low acetylation levels occurring at repressed genes. Twelve HDACs have been identified and can be sub-divided into three distinct

15   enzyme sub-families. These proteins are found *in vivo* as components of one or more co-repressor complexes, such as NCoR and Rb. They are also associated with cellular lifespan and ageing (Sir2), and are closely associated with cancer via their connections to Rb and p53 function, as well as via their association with complexes involved in enzymatic DNA methylation (DNMT1) and recognition of methylated DNA (MeCP2, NuRD).

20   HDACs appear to be involved in various disease states. For example, the majority of leukemias characterized to date conatin chromosomal translocations that aberrantly fuse DNA sequences encoding two different proteins. A significant proportion of these translocations occur between two different transcription factors; such translocations have been identified in multiple distinct leukemias. Strikingly, most of these appear to fuse repressor activity to a transcriptional

25   activator that, in the absence of repressor activity, activates genes involved in terminal differentiation. The aberrant chimeric repressors, resulting from translocation, instead repress these genes, in many cases by recruiting HDAC containing corepressor complexes such as NCoR and Sin3a. One example, APL, is the result of a PML-RARα translocation event, which splices a transcription factor that drives normal terminal differentiation (RARα) onto a potent repressor

30   (PML). High doses of the RARα ligand, trans-retinoic acid, can switch this chimera back to an activator and reduce the repressive activity of the aberrant APL. Co-treatment with a broad spectrum HDAC inhibitors seems to further reduce repressive activity of the APL, which appears to have therapeutic effects when given alone and in other leukemias. Other RARα translocations, such as PZLF-RARα, are not affected by trans-retinoic acid. Similarly, no modulators of the

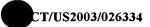35   activity of other translocations, such as AML1-ETO or TEL-AML1, have been identified to date.

C. DNA methyltransferases (DMTs)

DNA methyltransferases methylate C residues at CG doublets and at CpNpG sites, which occur preferentially near to/at gene regulatory elements. See, for example, WO 01/83793. High numbers/clusters of these CG doublets are called CpG islands and commonly occur at gene

5   promoters (perhaps as much as 50% of all genes have associated CpG islands). Unmethylated CGs occur in the vicinity of active genes, methyl-CpGs appear to be associated with silenced genes. This is of vital importance during transcriptional silencing and in development and is associated with transcriptional repression and silencing.

DNA methylation patterns are considerably altered in cancer and include genome-wide

10 losses of normal patterns of DNA methylation, together with specific increases in DNA methylation in the vicinity of tumor suppressor genes such as p16 and Rb. This oncogenic promoter hypermethylation and consequent transcriptional silencing at such genes has been shown to play important roles in tumorigenesis. Because of this linkage, inhibitors of DNA methyltransferases are of considerable interest as therapeutic compounds in the treatment of

15 cancer. Three functional human DNA methyltransferases have been identified, DNMT1, DNMT3a and DNMT3b, with the precise role of each at present undetermined. Complexes that recognize this methyl-C residues in DNA also been identified (MeCP2, NuRD/MeCP1). Some basic inhibitors exist which appear to inhibit the activity of all these enzymes, such as 5-azacytidine, and have already entered clinical trials. However, inhibitors specific to a particular DNA

20 methyltransferase or methylated DNA-recognizing protein are preferred, since they are prone to fewer side effects than are global inhibitors. The methods and compositions disclosed herein can be used to identify such inhibitors (as well as activators) by using one of the aforementioned proteins as a modifying activity in the disclosed fusion proteins.

25   ## D. Histone acetyltransferases (HATs)

Histone acetyltransferases acetylate lysine residues within histone tails as well as in other proteins such as p53. Typically, HATs are associated with transcriptional coactivators; with histone acetylation connected to active genes and gene activation *in vivo*, as well as in DNA damage repair. Exemplary HATs include p300, CBP, GCN5 and pCAF; additional HAT activities

30 and HAT-containing complexes are disclosed, for example, in WO 01/83793.

DNA-Binding Domains

In preferred embodiments, the compositions and methods disclosed herein involve use of DNA binding proteins, particular zinc finger proteins. A DNA-binding domain can comprise any

35 molecular entity capable of sequence-specific binding to chromosomal DNA. Binding can be mediated by electrostatic interactions, hydrophobic interactions, or any other type of chemical

23

interaction. Examples of moieties which can comprise part of a DNA-binding domain include, but are not limited to, minor groove binders, major groove binders, antibiotics, intercalating agents, peptides, polypeptides, oligonucleotides, and nucleic acids. An example of a DNA-binding nucleic acid is a triplex-forming oligonucleotide.

5   Minor groove binders include substances which, by virtue of their steric and/or electrostatic properties, interact preferentially with the minor groove of double-stranded nucleic acids. Certain minor groove binders exhibit a preference for particular sequence compositions. For instance, netropsin, distamycin and CC-1065 are examples of minor groove binders that bind specifically to AT-rich sequences, particularly runs of A or T. WO 96/32496.

10   Many antibiotics are known to exert their effects by binding to DNA. Binding of antibiotics to DNA is often sequence-specific or exhibits sequence preferences. Actinomycin, for instance, is a relatively GC-specific DNA binding agent.

In a preferred embodiment, a DNA-binding domain is a polypeptide. Certain peptide and polypeptide sequences bind to double-stranded DNA in a sequence-specific manner. For example,

15   transcription factors participate in transcription initiation by RNA Polymerase II through sequence-specific interactions with DNA in the promoter and/or enhancer regions of genes. Defined regions within the polypeptide sequence of various transcription factors have been shown to be responsible for sequence-specific binding to DNA. *See,* for example, Pabo *et al.* (1992) *Ann. Rev. Biochem.* 61:1053-1095 and references cited therein. These regions include, but are not limited to, motifs

20   known as leucine zippers, helix-loop-helix (HLH) domains, helix-turn-helix domains, zinc fingers, β-sheet motifs, steroid receptor motifs, bZIP domains homeodomains, AT-hooks and others. The amino acid sequences of these motifs are known and, in some cases, amino acids that are critical for sequence specificity have been identified. Polypeptides involved in other process involving DNA, such as replication, recombination and repair, will also have regions involved in specific

25   interactions with DNA. Peptide sequences involved in specific DNA recognition, such as those found in transcription factors, can be obtained through recombinant DNA cloning and expression techniques or by chemical synthesis, and can be attached to other components of a fusion molecule by methods known in the art.

In a more preferred embodiment, a DNA-binding domain comprises a zinc finger DNA-

30   binding domain. *See,* for example, Miller *et al.* (1985) *EMBO J.* 4:1609-1614; Rhodes *et al.* (1993) *Scientific American* Feb.:56-65; and Klug (1999) *J. Mol. Biol.* 293:215-218. The three-fingered Zif268 murine transcription factor has been particularly well studied. (Pavletich, N. P. & Pabo, C. O. (1991) *Science* 252:809-17). The X-ray co-crystal structure of Zif268 ZFP and double-stranded DNA indicates that each finger interacts independently with DNA (Nolte et al.

35   (1998) *Proc Natl Acad Sci U S A* 95:2938-43; Pavletich, N. P. & Pabo, C. O. (1993) *Science* 261:1701-7). The organization of the 3-fingered domain allows recognition of three contiguous
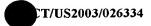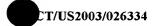
base-pair triplets by each finger. Each finger is approximately 30 amino acids long, adopting a
$\beta\beta\alpha$ fold. The two β-strands form a sheet, positioning the recognition α-helix in the major groove
for DNA binding. Specific contacts with the bases are mediated primarily by four amino acids
immediately preceeding and within the recognition helix. Conventionally, these recognition

5   residues are numbered -1, 2, 3, and 6 based on their positions in the α-helix.

ZFP DNA-binding domains are engineered (*e.g.*, designed and/or selected) to recognize a
particular target site as described in co-owned WO 00/42219; WO 00/41566 and WO 02/42459; as
well as U.S. Patents 5,789,538; 6,007,408; 6,013,453; 6,140,081; 6,140,466 and 6,242,568; and
PCT publications WO 95/19431, WO 98/53057, WO 98/53058, WO 98/53059, WO 98/53060,

10   WO 98/54311, WO 00/23464, WO 00/27878 and WO 01/53480. In one embodiment, a target site
for a zinc finger DNA-binding domain is identified according to site selection rules disclosed in co-
owned WO 00/42219. In certain embodiments, a ZFP is selected by iterative processes of selection
and optimization as described in co-owned International Patent Application PCT/US01/43568. In
additional embodiments, the binding specificity of the DNA-binding domain can be determined by

15   identifying accessible regions in the sequence in question (*e.g.*, in cellular chromatin). Accessible
regions can be determined as described in co-owned PCT publications WO 01/83732 and
WO 01/83751, the disclosures of which are hereby incorporated by reference herein. A DNA-
binding domain is then designed and/or selected as described herein to bind to a target site within
the accessible region.

20   Two alternative methods are typically used to create the coding sequences required to
express newly designed DNA-binding peptides. One protocol is a PCR-based assembly procedure
that utilizes six overlapping oligonucleotides. Three oligonucleotides correspond to "universal"
sequences that encode portions of the DNA-binding domain between the recognition helices.
These oligonucleotides remain constant for all zinc finger constructs. The other three "specific"

25   oligonucleotides are designed to encode the recognition helices. These oligonucleotides contain
substitutions primarily at positions -1, 2, 3 and 6 on the recognition helices making them specific
for each of the different DNA-binding domains.

The PCR synthesis is carried out in two steps. First, a double stranded DNA template is
created by combining the six oligonucleotides (three universal, three specific) in a four cycle PCR

30   reaction with a low temperature annealing step, thereby annealing the oligonucleotides to form a
DNA "scaffold." The gaps in the scaffold are filled in by high-fidelity thermostable polymerase,
the combination of Taq and Pfu polymerases also suffices. In the second phase of construction, the
zinc finger template is amplified by external primers designed to incorporate restriction sites at
either end for cloning into a shuttle vector or directly into an expression vector.

35   An alternative method of cloning the newly designed DNA-binding proteins relies on
annealing complementary oligonucleotides encoding the specific regions of the desired zinc finger

protein. This particular application requires that the oligonucleotides be phosphorylated prior to the final ligation step. Phosphorylation is usually performed before annealing, but can also be done post-annealing. In brief, the "universal" oligonucleotides encoding the constant regions of the proteins are annealed with their complementary oligonucleotides. Additionally, the "specific"

5   oligonucleotides encoding the finger recognition helices are annealed with their respective complementary oligonucleotides. These complementary oligos are designed to fill in the region, which was previously filled in by polymerase in the protocol described above. The complementary oligos to the common oligos 1 and finger 3 are engineered to leave overhanging sequences specific for the restriction sites used in cloning into the vector of choice. The second assembly protocol

10   differs from the initial protocol in the following aspects: the "scaffold" encoding the newly designed zinc finger protein is composed entirely of synthetic DNA thereby eliminating the polymerase fill-in step, additionally the fragment to be cloned into the vector does not require amplification. Lastly, inclusion in the design of sequence-specific overhangs eliminates the need for restriction enzyme digestion of the ZFP-encoding fragment prior to its insertion into the vector.

15   The resulting fragment encoding the newly designed zinc finger protein is ligated into an expression vector. Expression vectors that are commonly utilized include, but are not limited to, a modified pMAL-c2 bacterial expression vector (New England BioLabs, "NEB") or a eukaryotic expression vector, pcDNA (Promega). Conventional methods of purification can be used (*see* Ausubel, *supra*, Sambrook, *supra*). In addition, any suitable host can be used, e.g., bacterial cells,

20   insect cells, yeast cells, mammalian cells, and the like.

Expression of the zinc finger protein fused to a maltose binding protein (MBP-ZFP) in bacterial strain JM109 allows for straightforward purification through an amylose column (NEB). High expression levels of the zinc finger chimeric protein can be obtained by induction with IPTG since the MBP-ZFP fusion in the pMal-c2 expression plasmid is under the control of the IPTG

25   inducible tac promoter (NEB). Bacteria containing the MBP-ZFP fusion plasmids are inoculated in to 2x YT medium containing 10μM $ZnCl_2$, 0.02% glucose, plus 50 μg/ml ampicillin and shaken at 37°C. At mid-exponential growth IPTG is added to 0.3 mM and the cultures are allowed to shake. After 3 hours the bacteria are harvested by centrifugation, disrupted by sonication, and then insoluble material is removed by centrifugation. The MBP-ZFP proteins are captured on an

30   amylose-bound resin, washed extensively with buffer containing 20 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5 mM DTT and 50 μM $ZnCl_2$, then eluted with maltose in essentially the same buffer (purification is based on a standard protocol from NEB). Purified proteins are quantitated and stored for biochemical analysis.

The biochemical properties of the purified proteins, e.g., K_d, can be characterized by any

35   suitable assay. K_d can be characterized via electrophoretic mobility shift assays ("EMSA") (Buratowski & Chodosh, *in Current Protocols in Molecular Biology* pp. 12.2.1-12.2.7 (Ausubel
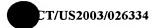
ed., 1996); *see also* U.S. Patent No. 5,789,538, and PCT WO 00/42219, herein incorporated by

reference). Affinity is measured by titrating purified protein against a low fixed amount of labeled

double-stranded oligonucleotide target. The target comprises the natural binding site sequence

(*e.g.*, 9 or 18 bp), optionally flanked by the 3 bp found in the natural sequence. External to the

5   binding site plus flanking sequence is a constant sequence. The annealed oligonucleotide targets

possess a 1-nucleotide 5' overhang that allows for efficient labeling of the target with T4 phage

polynucleotide kinase. For the assay the target is added at a concentration of 40 nM or lower (the

actual concentration is kept at least 10-fold lower than the lowest protein dilution) and the reaction

is allowed to equilibrate for at least 45 min. In addition the reaction mixture also contains 10 mM

10   Tris (pH 7.5), 100 mM KCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 5 mM DTT, 10% glycerol, 0.02% BSA

(poly (dIdC) or (dAdT) (Pharmacia) can also added at 10-100 µg/µl).

The equilibrated reactions are loaded onto a 10% polyacrylamide gel, which has been pre-

run for 45 min in Tris/glycine buffer, then bound and unbound labeled target is resolved be

electrophoresis at 150V (alternatively, 10-20% gradient Tris-HCl gels, containing a 4%

15   polyacrylamide stacker, can be used). The dried gels are visualized by autoradiography or

phosphoroimaging and the apparent K_d is determined by calculating the protein concentration that

gives half-maximal binding.

Similar assays can also include determining active fractions in the protein preparations.

Active fractions are determined by stoichiometric gel shifts where proteins are titrated against a

20   high concentration of target DNA. Titrations are done at 100, 50, and 25% of target (usually at

micromolar levels).

Fusion Molecules

In the compositions and methods described herein, zinc finger-containing proteins that

25   target specific sequences are generally provided as fusion molecules in combination with other

molecules, particularly with one or more modifying domains (*e.g.*, CMEs). Thus, in certain

embodiments, the compositions and methods disclosed herein involve fusions between at least one

of the zinc finger proteins described herein (or functional fragments thereof) and one or more

CMEs (or functional fragments thereof), or a polynucleotide encoding such a fusion. In such a

30   fusion molecule, the CME is brought into proximity with a sequence in a gene that is bound by the

zinc finger protein and can function, for example, by regulating expression of the gene. Changes in

regulation of a target gene by a fusion protein provides an assay for molecules that modulate the

activity of a CME.

The zinc finger protein can be covalently or non-covalently associated with one or more

35   regulatory domains (*e.g.*, CMEs), alternatively two or more regulatory domains, with the two or

more domains being two copies of the same domain, or two different domains. The regulatory

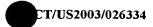27

domains can be covalently linked to the zinc finger protein, e.g., via an amino acid linker, as part of a fusion protein. The zinc finger proteins can also be associated with a regulatory domain via a non-covalent dimerization domain, e.g., a leucine zipper, a STAT protein N terminal domain, or a protein which binds cyclosporine, tetracycline, a steroid, FK506, FK520, rapamycin, and analogues

5   or derivatives thereof. Examples of such proteins include FK506 binding proteins (FKBPs), cyclophilin receptors, tetracycline receptors, steroid receptors and FRAPs. *See, e.g.,* US Patent No. 6,165,787; O'Shea, *Science* 254: 539 (1991), Barahmand-Pour *et al., Curr. Top. Microbiol. Immunol.* 211:121-128 (1996); Klemm *et al., Annu. Rev. Immunol.* 16:569-592 (1998); Ho *et al., Nature* 382:822-826 (1996); and Pomeranz *et al., Biochem.* 37:965 (1998). The regulatory domain

10   can be associated with the zinc finger protein at any suitable position, including the C- or N-terminus of the zinc finger protein.

A. Construction of Fusion Molecules

Fusion molecules may be constructed by methods of cloning and biochemical conjugation that are well known to those of skill in the art. In certain embodiments, fusion molecules comprise

15   a zinc finger protein and one or more CMEs (or functional fragments thereof). Optionally, fusion molecules also comprise nuclear localization signals (such as, for example, that from an SV40 T-antigen) and epitope tags (such as, for example, FLAG, *myc* and hemagglutinin). Fusion proteins (and nucleic acids encoding them) are designed such that the translational reading frame is

20   preserved among the components of the fusion.

The fusion molecules disclosed herein comprise a zinc finger binding protein that binds to a target site (in a reporter gene) and a CME. Preferably, the target site is in an endogenous gene whose level of expression can be readily assayed. Modulation of gene expression can be in the form of increased expression or repression. The effect of a compound or substance on the

25   regulation of the reporter gene by the fusion protein can then be assayed to determine if a compound or substance is a modulator of the CME.

For any such applications, the fusion molecule(s) can be formulated with a pharmaceutically acceptable carrier, as is known to those of skill in the art. See, for example, Remington's Pharmaceutical Sciences, 17th ed., 1985; and co-owned WO 00/42219.

30   Linker domains between polypeptide domains, e.g., between the zinc finger proteins and the CME, can be included. Such linkers are typically polypeptide sequences, such as poly gly sequences of between about 5 and 200 amino acids. Preferred linkers are typically flexible amino acid subsequences that are synthesized as part of a recombinant fusion protein, for example, the linkers DGGGS (SEQ ID NO:5); TGEKP (SEQ ID NO:6) (*see, e.g.,* Liu *et al., Proc. Natl. Acad.*

35   *Sci. U.S.A.* 5525-5530 (1997)); LRQKDGERP (SEQ ID NO:7); GGRR (SEQ ID NO:8) (Pomerantz *et al.* 1995, *supra*); (G4S)$_n$ (SEQ ID NO:9) (Kim *et al., Proc. Natl. Acad. Sci. U.S.A.*
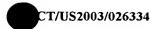
28

93, 1156-1160 (1996); GGRRGGGS (SEQ ID NO:10); LRQRDGERP (SEQ ID NO:11);
LRQKDGGGSERP (SEQ ID NO:12); and LRQKd(G3S)$_2$ ERP (SEQ ID NO:13). Additional
suitable linkers are disclosed in WO 99/45132 and WO 01/53480.

     A chemical linker can be used to connect synthetically or recombinantly produced domain

5   sequences. For example, poly(ethylene glycol) linkers are available from Shearwater Polymers,
Inc. Huntsville, Alabama. Some linkers have amide linkages, sulfhydryl linkages, or
heterofunctional linkages. In addition to covalent linkage of zinc finger proteins to regulatory
domains, non-covalent methods can be used to produce molecules with zinc finger proteins
associated with regulatory domains. See, for example, US Patent No. 6,165,787 and

10   WO 01/30843.

     As noted above, the fusion molecules may be in the form of nucleic acid sequences that
encode the fusion molecule ,or in the form of a fusion between one or more polypeptides and/or
one or more polypeptides and one or more non-polypeptide molecules.

15   <u>B. Additional Functional Domains</u>

     In addition to the CME, the fusion molecule may also include one or more additional
regulatory (functional) domains including, *e.g.*, effector domains from transcription factors
(activators, repressors, co-activators, co-repressors), silencers, nuclear hormone receptors,
oncogene transcription factors (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family

20   members etc.); DNA repair enzymes and their associated factors and modifiers; DNA
rearrangement enzymes and their associated factors and modifiers; chromatin associated proteins
and their modifiers (e.g., kinases, acetylases, deacetylases, phosphatases, methyltransferases,
ubiquitinylases); and DNA modifying enzymes (e.g., methyltransferases, topoisomerases,
helicases, ligases, kinases, phosphatases, polymerases, and/or endonucleases, and their associated

25   factors and modifiers.

     Transcription factor polypeptides from which one can obtain a regulatory domain include
those that are involved in regulated and basal transcription. Such polypeptides include
transcription factors, their effector domains, coactivators, silencers, nuclear hormone receptors
(*see, e.g.*, Goodrich *et al.*, *Cell* 84:825-30 (1996) for a review of proteins and nucleic acid elements

30   involved in transcription; transcription factors in general are reviewed in Barnes & Adcock, *Clin.
Exp. Allergy* 25 Suppl. 2:46-9 (1995) and Roeder, *Methods Enzymol.* 273:165-71 (1996)).
Databases dedicated to transcription factors are known (*see, e.g., Science* 269:630 (1995)).
Nuclear hormone receptor transcription factors are described in, for example, Rosen *et al., J. Med.
Chem.* 38:4855-74 (1995). The C/EBP family of transcription factors are reviewed in Wedel *et al.*,

35   *Immunobiology* 193:171-85 (1995). Coactivators and co-repressors that mediate transcription
regulation by nuclear hormone receptors are reviewed in, for example, Meier, *Eur. J. Endocrinol.*
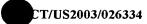
134(2):158-9 (1996); Kaiser *et al.*, *Trends Biochem. Sci.* 21:342-5 (1996); and Utley *et al.*, *Nature* 394:498-502 (1998)). GATA transcription factors, which are involved in regulation of hematopoiesis, are described in, for example, Simon, *Nat. Genet.* 11:9-11 (1995); Weiss *et al.*, *Exp. Hematol.* 23:99-107. TATA box binding protein (TBP) and its associated TAF polypeptides

5   (which include TAF30, TAF55, TAF80, TAF110, TAF150, and TAF250) are described in Goodrich & Tjian, *Curr. Opin. Cell Biol.* 6:403-9 (1994) and Hurley, *Curr. Opin. Struct. Biol.* 6:69-75 (1996). The STAT family of transcription factors are reviewed in, for example, Barahmand-Pour *et al.*, *Curr. Top. Microbiol. Immunol.* 211:121-8 (1996). Transcription factors involved in disease are reviewed in Aso *et al.*, *J. Clin. Invest.* 97:1561-9 (1996).

10   An exemplary functional domain for fusing with a ZFP is a KRAB repression domain from the human KOX-1 protein (see, e.g., Thiesen et al., New Biologist 2, 363-374 (1990); Margolin et al., Proc. Natl. Acad. Sci. USA 91, 4509-4513 (1994); Pengue et al., Nucl. Acids Res. 22:2908-2914 (1994); Witzgall et al., Proc. Natl. Acad. Sci. USA 91, 4514-4518 (1994). Another suitable repression domain is methyl binding domain protein 2B (MBD-2B) (see, also Hendrich et al.

15   (1999) *Mamm Genome* 10:906-912 for description of MBD proteins). Another useful repression domain is that associated with the v-ErbA protein. *See*, for example, Damm, et al. (1989) *Nature* 339:593-597; Evans (1989) *Int. J. Cancer Suppl.* 4:26-28; Pain et al. (1990) *New Biol.* 2:284-294; Sap et al. (1989) *Nature* 340:242-244; Zenke et al. (1988) *Cell* 52:107-119; and Zenke et al. (1990) *Cell* 61:1035-1049. Additional exemplary repression domains include, but are not limited

20   to, thyroid hormone receptor (TR), SID, MBD2, MBD3, members of the DNMT family (*e.g.*, DNMT1, DNMT3A, DNMT3B), Rb, and MeCP2. *See*, for example, Zhang et al. (2000); *Ann Rev Physiol* 62:439-466; Bird *et al.* (1999) *Cell* 99:451-454; Tyler *et al.* (1999) *Cell* 99:443-446; Knoepfler *et al.* (1999) *Cell* 99:447-450; and Robertson *et al.* (2000) *Nature Genet.* 25:338-342. Additional exemplary repression domains include, but are not limited to, ROM2 and AtHD2A.

25   *See*, for example, Chern *et al.* (1996) *Plant Cell* 8:305-321; and Wu *et al.* (2000) *Plant J.* 22:19-27.

Suitable domains for achieving activation include the HSV VP16 activation domain (see, e.g., Hagmann et al., J. Virol. 71, 5952-5962 (1997)) nuclear hormone receptors (see, e.g., Torchia et al., Curr. Opin. Cell. Biol. 10:373-383 (1998)); the p65 subunit of nuclear factor kappa B (Bitko

30   & Barik, J. Virol. 72:5610-5618 (1998)and Doyle & Hunt, Neuroreport 8:2937-2942 (1997)); Liu et al., Cancer Gene Ther. 5:3-28 (1998)), or artificial chimeric functional domains such as VP64 (Seifpal et al., EMBO J. 11, 4961-4968 (1992)).

Additional exemplary activation domains include, but are not limited to, VP16, VP64, p300, CBP, PCAF,SRC1 PvALF, AtHD2A and ERF-2. *See*, for example, Robyr *et al.* (2000) *Mol.*

35   *Endocrinol.* 14:329-347; Collingwood *et al.* (1999) *J. Mol. Endocrinol.* 23:255-275; Leo *et al.* (2000) *Gene* 245:1-11; Manteuffel-Cymborowska (1999) *Acta Biochim. Pol.* 46:77-89; McKenna
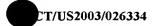
et al. (1999) *J. Steroid Biochem. Mol. Biol.* 69:3-12; Malik *et al.* (2000) *Trends Biochem. Sci.* 25:277-283; and Lemon *et al.* (1999) *Curr. Opin. Genet. Dev.* 9:499-504. Additional exemplary activation domains include, but are not limited to, OsGAI, HALF-1, C1, AP1, ARF-5, -6, -7, and − 8, CPRF1, CPRF4, MYC-RP/GP, and TRAB1. *See*, for example, Ogawa *et al.* (2000) *Gene*

5   245:21-29; Okanami *et al.* (1996) *Genes Cells* 1:87-99; Goff *et al.* (1991) *Genes Dev.* 5:298-309; Cho *et al.* (1999) *Plant Mol. Biol.* 40:419-429; Ulmason *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96:5844-5849; Sprenger-Haussels *et al.* (2000) *Plant J.* 22:1-8; Gong *et al.* (1999) *Plant Mol. Biol.* 41:33-44; and Hobo *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96:15,348-15,353.

Additional functional domains are disclosed, for example, in co-owned WO 00/41566.

10   Further, insulator domains, chromatin remodeling proteins such as ISWI-containing domains and/or methyl binding domain proteins suitable for use in fusion molecules are described, for example, in co-owned International Publications WO 02/26959; WO 02/26960; and WO 02/44376.

Kinases, phosphatases, and other proteins that modify polypeptides involved in gene regulation are also useful as functional domains for zinc finger proteins. Such modifiers are often

15   involved in switching on or off transcription mediated by, for example, hormones. Kinases involved in transcription regulation are reviewed in Davis, *Mol. Reprod. Dev.* 42:459-67 (1995), Jackson *et al.*, *Adv. Second Messenger Phosphoprotein Res.* 28:279-86 (1993), and Boulikas, *Crit. Rev. Eukaryot. Gene Expr.* 5:1-77 (1995), while phosphatases are reviewed in, for example, Schonthal & Semin, *Cancer Biol.* 6:239-48 (1995). Nuclear tyrosine kinases are described in

20   Wang, *Trends Biochem. Sci.* 19:373-6 (1994).

Useful domains can also be obtained from the gene products of oncogenes (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family members) and their associated factors and modifiers. Oncogenes are described in, for example, Cooper, *Oncogenes, The Jones and Bartlett Series in Biology* (2nd ed., 1995). The ets transcription factors are reviewed in Waslylk *et al.*, *Eur.*

25   *J. Biochem.* 211:7-18 (1993) and Crepieux *et al.*, *Crit. Rev. Oncog.* 5:615-38 (1994). Myc oncogenes are reviewed in, for example, Ryan *et al.*, *Biochem. J.* 314:713-21 (1996). The jun and fos transcription factors are described in, for example, *The Fos and Jun Families of Transcription Factors* (Angel & Herrlich, eds. 1994). The max oncogene is reviewed in Hurlin *et al.*, *Cold Spring Harb. Symp. Quant. Biol.* 59:109-16. The myb gene family is reviewed in Kanei-Ishii *et*

30   *al.*, *Curr. Top. Microbiol. Immunol.* 211:89-98 (1996). The mos family is reviewed in Yew *et al.*, *Curr. Opin. Genet. Dev.* 3:19-25 (1993).

Zinc finger proteins can include functional domains obtained from DNA repair enzymes and their associated factors and modifiers. DNA repair systems are reviewed in, for example, Vos, *Curr. Opin. Cell Biol.* 4:385-95 (1992); Sancar, *Ann. Rev. Genet.* 29:69-105 (1995); Lehmann,

35   *Genet. Eng.* 17:1-19 (1995); and Wood, *Ann. Rev. Biochem.* 65:135-67 (1996). DNA rearrangement enzymes and their associated factors and modifiers can also be used as regulatory
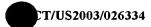
domains (see, e.g., Gangloff et al., Experientia 50:261-9 (1994); Sadowski, FASEB J. 7:760-7 (1993)).

In addition to functional domains, often the zinc finger protein is expressed as a fusion protein such as maltose binding protein ("MBP"), glutathione S transferase (GST), hexahistidine, c-myc, and the FLAG epitope, for ease of purification, monitoring expression, or monitoring cellular and subcellular localization.

Reporters

Virtually any component of a cell can serve as a molecular target (reporter) for the ZFP component of the fusion protein. For example, the product (mRNA or protein) of an endogenous cellular genes such as, e.g., VEGF, H19 or IGF-2, can serve as reporter. A gene whose product is used as a reporter is denoted a "reporter gene." An exogenous gene can also serve as a reporter gene, for example, if it is integrated into the chromosome so that it adopts a chromatin configuration. Additional non-limiting examples of endogenous reporters include growth factor receptors (e.g., FGFR, PDGFR, EGFR, NGFR, and VEGFR). Other endogenous reporters are G-protein receptors and include substance K receptor, the angiotensin receptor, the α- and β-adrenergic receptors, the serotonin receptors, and PAF receptor. See, e.g., Gilman, Ann. Rev. Biochem. 56:625-649 (1987). Other suitable reporters that may be employed include ion channels (e.g., calcium, sodium, potassium channels), muscarinic receptors, acetylcholine receptors, GABA receptors, glutamate receptors, and dopamine receptors (see Harpold, 5,401,629 and US 5,436,128). Other targets are adhesion proteins such as integrins, selectins, and immunoglobulin superfamily members (see Springer, Nature 346:425-433 (1990). Osborn (199) Cell 62:3; Hynes (1992) Cell 69:11). Other endogenous reporters are cytokines, such as interleukins IL-1 through IL-13, tumor necrosis factors α & β, interferons α, β and γ, tumor growth factor Beta (TGF-β), colony stimulating factor (CSF) and granulocyte monocyte colony stimulating factor (GM-CSF). See Human Cytokines: Handbook for Basic & Clinical Research (Aggrawal et al. eds., Blackwell Scientific, Boston, MA 1991). Other targets are hormones, enzymes, and intracellular and intercellular messengers, such as, adenyl cyclase, guanyl cyclase, and phospholipase C, nuclear receptors (e.g., FXR (Farnesoid X Receptor), PPARb (Peroxisome Proliferator Activator Receptor Delta), and RZR (Retinoid Z Receptor)), and organelle receptors. Target molecules that serve as reporter molecules can be human, mammalian viral, plant, fungal or bacterial. Other targets are antigens, such as proteins, glycoproteins and carbohydrates from microbial pathogens, both viral and bacterial, and tumors. Still other targets are described in U.S. Patent No. 4,366,241.

Reporter expression can be directly detected by detecting formation of transcript or of translation product. For example, transcription product can be detected using Northern blots, branched DNA signal amplification systems (e.g., US Patent Nos. 5,124,246; 5,624,802;
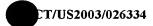
32

5,635,352; 5,681,697; 5,849,481), RNA tags (Aclara Biosciences, Mountain View, CA) or real-time PCR (Taqman®, Roche) and the formation of certain proteins can be detected, *e.g.*, by gel electrophoresis, immunoassay (*e.g.*, ELISA), using a characteristic stain or by detecting an inherent characteristic (*e.g.*, enzymatic activity) of the protein. Additionally, expression of reporter is

5   determined by detecting a product formed as a consequence of an activity of the reporter.

Exemplary reporter genes encoding proteins having enzymatic activity include, but are not limited to, those encoding phosphatases, hydrolases, myeloperoxidases and proteases. Additional exemplary reporter genes include those encoding cell-surface proteins such as, for example, CD antigens, immunoglobulins, T-cell receptors, growth factor receptors and transmembrane proteins

10   (*e.g.*, placental alkaline phosphatase).

Thus, cell based screening assays can be performed in which ZFP-CME fusions are introduced into a cell and used to identify compounds that inhibit or activate the CME(s). Readout is provided by the product of a reporter gene, the reporter gene being targeted by the ZFP portion of the chimeric protein. In certain embodiments, the ZFP-CME is constitutively expressed and the

15   effect of a compound on expression of the reporter gene is tested and compared to a baseline expression level prior to administration of the compound. In other embodiments, expression of the ZFP-CME is controlled by an inducible promoter and the ability of a compound to modulate CME activity is tested in the presence of inducer (and compared to values obtained in the presence of inducer and absence of compound). If the compound under test targets an intracellular component

20   other than the CME of interest, effects of the compound in the absence of inducer will be detected. The compound can be administered directly into a cell using methods known in the art and described herein.

In addition, any of the methods described herein can be used with any reporter and/or selectable marker. Reporters that can be directly detected include GFP (green fluorescent protein).

25   Fluorescence generated from this protein can be detected using a variety of commercially available fluorescent detection systems, including a FACS system for example. Other reporters are enzymes that catalyze the formation of a detectable product. Suitable enzymes include proteases, nucleases, liposes, phosphatases, sugar hydrolases and esterases. Preferably, the substrate is substantially impermeable to eukaryotic plasma membranes, thus making it possible to tightly control signal

30   formation. Examples of suitable reporter genes that encode enzymes include, for example, CAT (chloramphenicol acetyl transferase; Alton and Vapnek (1979) Nature 282:864-869), luciferase (lux), β-galactosidase, β-glucuronidase (GUS) and alkaline phosphatase (Toh, et al. (1980) Eur. J. Biochem. 182:231-238; and Hall et al. (1983) J. Mol. Appl. Gen. 2:101).

Selectable markers can also be used instead of, or in addition to, reporters. Positive

35   selection markers are those polynucleotides that encode a product that enables only cells that carry and express the gene to survive and/or grow under certain conditions. For example, cells that

33

express neomycin resistance (Neo^r) gene are resistant to the compound G418, while cells that do not express Neo^r are killed by G418. Other examples of positive selection markers including hygromycin resistance and the like will be known to those of skill in the art. Negative selection markers are those polynucleotides that encode a produce that enables only cells that carry and

5   express the gene to be killed under certain conditions. For example, cells that express thymidine kinase (*e.g.,* herpes simplex virus thymidine kinase, HSV-TK) are killed when gancyclovir is added. Other negative selection markers are known to those skilled in the art. The selectable marker need not be a transgene and, additionally, reporters and selectable markers can be used in various combinations.

10

Screening Assays

The emphasis of pharmaceutical research activities has shifted toward the purposeful discovery of novel chemical classes and novel molecular targets. This change in emphasis, and timely technological breakthroughs (*e.g.,* molecular biology, laboratory automation, combinatorial

15   chemistry) gave birth to high throughput screening, or HTS, which is now widespread throughout the biopharmaceutical industry.

High throughput screening involves several steps: creating an assay that is predictive of a particular physiological response; automating the assay so that it can be reproducibly performed a large number of times; and, sequentially testing samples from a chemical library to identify

20   chemical structures able to "hit" the assay, suggesting that such structures might be capable of provoking the intended physiological response. Hits from the high throughput screen are followed up in a variety of secondary assays to eliminate artifactual results, particularly toxic compounds.

Thus, the assays used in high throughput screens are intended to detect the presence of chemical samples (*e.g.,* compounds, substances, molecules) possessing specific biological or

25   biochemical properties. These properties are chosen to identify compounds with the potential to elicit a specific biological response when applied *in vivo*. High throughput screens identify both agents that can be used as drugs themselves and, in addition, drug candidates that will ultimately be used as drugs. A compound of a certain chemical class that is found to have some level of desired biological property in a high-throughput assay can then be the basis for synthesis of derivative

30   compounds.

Assays generally fall into two broad categories: biochemical assays and cell-based assays. Biochemical assays utilize pure or semi-pure components outside of a cellular environment. Enzyme assays and receptor binding assays are typical examples of biochemical assays. Cell-based assays utilize intact cells in culture. Examples of such assays include luciferase reporter

35   gene assays and calcium flux assays.

Biochemical assays are usually easier to perform and are generally less prone to artifacts than conventional cell-based assays. Compounds identified as "active" in a biochemical assay typically function according to a desired mechanism, decreasing the amount of follow-up experimentation required to confirm a compound's status as a "hit." A major disadvantage of

5   biochemical assays, however, is the lack of biological context. Compound "hits" from biochemical screens do not have to traverse a plasma membrane or other structures to reach and affect the target protein. Consequently, biochemical assays tend to be far less predictive of a compound's activity in an animal than cell-based assays.

Cell-based assays preserve much of the biological context of a molecular target.

10   Compounds that cannot pass through the plasma membrane or that are toxic to the cell are not pursued. This context, however, adds complexity to the assay. Therefore conventional cell-based assays are much more prone to artifact or false positive results than are biochemical assays. Compounds that trigger complex toxic reactions or trigger apoptosis are particularly troublesome. Much of the labor devoted to conventional cell-based high throughput screening is directed to

15   follow-up assays that detect false hits or hits that work by undesirable mechanisms.

If false positive or artifactual hits could be rapidly identified and eliminated, the ease and efficiency of biochemical assays could be approached in cell-based assays, while preserving the biological context. The result would be an assay with optimum throughput and optimum predictability of biological function. In short, a more efficient process for the discovery of new

20   pharmaceuticals would be produced.

As disclosed herein, screening assays are conducted using cells comprising at least one targeted DNA-binding domain (e.g., an engineered zinc finger protein) and at least one CME. The screening assays described herein allow for high throughput screening of candidate compounds and can be accomplished while reducing false positives. As a result, discovery will be more efficient

25   and compounds identified using the screening methods disclosed herein will have greater specificity and, consequently, will be prone to fewer potential side effects than those identified by prior methods.

A. Cell-Based Assays

30   In one aspect, methods of performing cell-based assays that identify modulators of CME activity are described. For example, cell-based assays as disclosed herein can utilize a cell line which expresses (either constitutively or inducibly) a chimeric protein comprising a fusion of the catalytic domain of CME with a targeted DNA binding domain (e.g., an engineered ZFP) that binds one or more reporter genes. The fusion protein(s) are able to up- or down-regulate transcription of

35   reporter gene(s), which can be either endogenous or exogenous. The cells can be transiently or
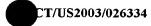
35

stably transfected with a polynucleotide encoding the fusion molecule. In this way, a direct, high-throughput screen for compounds that regulate CME activity is created.

     In certain embodiments, a ZFP-CME fusion represses transcription of a reporter gene to which the ZFP is targeted, in cells in which the fusion protein is expressed. Compounds which,

5   when contacted with such cells, result in increased reporter gene expression (*i.e.*, relief of repression) are modulators of the CME.

     Some cells are designed to express a sequence encoding a ZFP-CME fusion protein in operable linkage to an inducible promoter. A variety of inducible promoters are available that can be regulated by small molecules or other stimuli such as heat. Operable linkage to an inducible

10   promoter allows activation of a ZFP-CME fusion, and thereby modulation of a molecular target by the expressed fusion protein, to be controlled by supplying the cell with the appropriate small molecule or other stimulus. Use of inducible promoters is useful for achieving transient modulation of cellular proteins whose permanent over- or under-expression would result in lethality to the cell. Inducible expression is also advantageous in reducing secondary effects due to

15   modulation of an intended cellular protein. For example, regulation of one cellular protein can directly or indirectly results in changes in the relative abundance of many others proteins within the cell. By inducing a ZFP-CME fusion protein shortly before an assay is performed, such secondary changes are minimized. Accordingly, differences in response between a test cell and a control cell having a molecular target or other protein subject to regulation are entirely or substantially entirely

20   due to interaction between the compound and the molecular target rather than secondary effects caused by regulation of the target.

     The present methods and compositions allow for a direct transcriptional readout of the activities of chromatin modifying catalytic activities *in vivo*. This in turn, allows high-throughput screening for potential inhibitors and activators of these activities. The direct targeting of

25   chromatin (*e.g.*, histone) modifications to a particular reporter gene, via fusions of the catalytic domain of a CME to a targeted DNA-binding domain (*e.g.*, an engineered ZFP) results, in certain embodiments, in transcriptional regulation an endogenous gene. Changes in such targeted gene regulation can be utilized as an activity assay at the RNA or protein level, providing a direct, rapid assay system for use in screening. In certain embodiments, expression of a ZFP-CME fusion is

30   regulatable, for example by an inducible system such as TREx (Invitrogen, Carlsbad, CA; see also US Patent No. 4,833,080), allowing for specific control of the expression of ZFP-CME fusion proteins. In the case of the aforementioned T-REx system, transcription of a polynucleotide sequence encoding a ZFP-CME fusion is positively regulated by provision of tetracycline or a tetracycline analogue, such as, for example, doxycycline.

35   Using the compositions and methods described herein, modulators of individual enzymes such as, for example, modulators of enzymes in the histone H3 lysine 9 methyltransferases G9A,
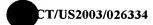
SETDB1, Suv39H1 and H2, histone H3 Lysine 4, 27 or 36 methyltransferases, and H4 lysine 20 methyltransferases can be identified. Fusion proteins can also be designed and synthesized to assay for inhibitors of histone acetyltransferases, histone deacetylases, arginine methyltransferases such as CARM1 or PRMT1, or any transcriptional modulator which functions via modification of

5 chromatin structure. In addition, the methods and compositions disclosed herein allow for screening of *in vivo* modulators (*e.g.*, activators, inhibitors) of specific DNA methyltransferases such as DNMT1, DNMT3A and DMNT3b.

Moreover, the cells and cell lines described herein are suitable for the subsequent confirmation of the action of modulators identified via conventional methodologies (*e.g.*,

10 immunoprecipitation), further enhancing the functionality of these cell lines and streamlining the validation of potential modulators, and strengthening the utility of the disclosed methods and compositions.

Cells comprising fusions of ZFPs and CMEs may be generated by any method known in the art. For example, a cell or cell line can be stably transfected or transiently transfected with

15 nucleic acids encoding the fusions.

In certain embodiments, test and control cells (or cell lines) are used in which the test cell is substantially identical (*e.g.*, isogenic) to the control cell except for the presence of the fusion molecule (and, possibly, a low incidence of random mutations resulting from environmental factors). An "isogenic cell" is a cell that, with the possible exception of a few random mutations

20 due to environmental factors, contains identical genetic material to that of another cell. Accordingly, typically >99% or 99.9 or 99.99% of the genetic material of one cell is identical to that of another. Thus, in certain embodiments, the phenotype of the test and controls cell populations will differ only in regard to the levels of the fusion protein(s). In other embodiments, the test and control cells will differ only with respect to the compounds to which they are exposed

25 during testing. In additional embodiments, test and control cells will both comprise similar or identical levels of a ZFP-CME fusion protein, but will differ in that the test cells are exposed to a compound and the control cells are not.

The cells can be individual cells or a population, the latter being more usual. The cell types can be cell lines or natural (*e.g.*, isolated) cells. Cell lines are available, for example from the

30 American Type Culture Collection (ATCC), or can be generated by methods known in the art, as described for example in Sambrook et al., *supra*. Similarly cells can be isolated by methods known in the art. Other non-limiting examples of cell types include cells that have or are subject to pathologies, such as cancerous cells, or pathogenically infected cells; stem cells; fully differentiated cells; partially differentiated cells; immortalized cells and the like. Both prokaryotic

35 (*e.g.*, bacteria) and eukaryotic (*e.g.*, yeast, plant, insect, fungal, piscine and mammalian cells such as feline, canine, murine, bovine, equine, caprine, porcine, ovine, primate and human) cells can be
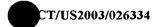
used with eukaryotic cells being preferred. Mammalian (human and non-human) cell types are particularly preferred. The choice of cell type depends in part on the intended recipient of a drug. For example, human cell types are advantageous for screening drugs intended for use in human, and feline cell types are advantageous for screening drugs intended for use in cats.

5   Suitable mammalian cells include CHO (chinese hamster ovary) cells, HEP-G2 cells, BaF-3 cells, Schneider cells, COS cells (monkey kidney cells expressing SV40 T-antigen), CV-1 cells, HuTu80 cells, NTERA2 cells, NB4 cells, HL-60 cells, HeLa cells, MCF-7 cells, U2OS cells, 293 cells (see, e.g., Graham et al. (1977) J. Gen. Virol. 36:59), and myeloma cells like SP2 or NS0 (see, e.g., Galfre and Milstein (1981) Meth. Enzymol. 73(B):3-46).
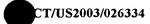
10   Other eukaryotic cells include, for example, insect (e.g., sp. *frugiperda*), yeast (e.g., *S. cerevisiae, S. pombe, P. pastoris, K. lactis, H. polymorpha*) and fungal and plant cells (Fleer, R. (1992) Current Opinion in Biotechnology 3:486-496). Bacterial cell types include *E. coli, B. subtilis* and *S. typhimurium.*

Cells can be transiently or stably transfected or transformed with a ZFP-CME fusion
15   molecule (or polynucleotide encoding the fusion molecule). Methods of transfecting cells are known in the art and described, for example, in Ausubel et al, *supra.*

One can use a number of methods in applying cell-based assays for discovery of CME modulators. For example, expression of a reporter gene is measured in a control cell population expressing a ZFP-CME fusion. Expression of the ZFP-CME fusion is preferably inducible, *e.g.*,
20   upon provision of a small molecule such as tetracycline or doxycycline. The value of reporter gene expression serves as a baseline measure for the assay. One then contacts a test cell population, also expressing the ZFP-CME fusion, with a candidate compound and measures reporter gene expression. A statistically different value for the two responses indicates that the compound is a "hit"; it substantially interacts with, and modulates the activity of, the CME. The measurements
25   can also be conducted in the opposite order, in which one first assays the test cell population.

In another method, two different cell populations can, for example, be put in two different vessels. A solution of a candidate compound can be added sequentially to both vessels and transcription of a reporter gene determined. Transcription can be determined in any number of ways, for example, by measuring mRNA levels and/or expression of the protein using techniques
30   known in the art and described herein. When there is a significant difference (*i.e.*, outside the scope of experimental error) between values of reporter gene expression for the respective cell populations, one determines the candidate compound to be a "hit" in the assay.

In another method, which can be used to assay for non-specific effects of a compound, a control cell population expresses a ZFP-CME fusion but is not contacted with a compound. A first
35   test cell population also expresses the ZFP-CME fusion and is contacted with the compound. A second test cell population, not expressing the ZFP-CME fusion, is also contacted with the

compound. The first test cell population is compared with the control population to determine whether the compound modulates the regulation of the reporter gene by the ZFP-CME fusion. If an effect of the compound on the reporter gene regulation is observed, expression of the reporter gene in the second test cell population is examined. An effect of the compound on reporter gene

5   expression in the second test cell population is an indication of non-specificity of the compound and would therefore disqualify the compound as a hit. In certain embodiments, in which expression of the ZFP-CME fusion is inducible, control cells and the first test cell population is exposed to inducer and the second test cell population is not exposed to inducer.

In some methods, analysis of cellular response in test and control cells is performed in

10   parallel. In other methods, analysis of cellular response in test cells is compared with historical controls. In some methods, cellular response in presence of a compound in either test or control cells is compared with the response of like cells in absence of the compound.

In some methods, compounds are screened individually. In other methods, many compounds are screened in parallel. Microtiter plates and robotics are particularly useful for

15   parallel screening of many compounds. Optical detection can be employed for rapidity and automation. Hundreds, thousands or even millions of compounds can be screened per week.

Once a "hit" is identified using the present methods, derivatives of the compound can be made to maximize its ability to interact with its molecular target. Derivatives can be produced using conventional techniques such as self-consistent field (SCF) analysis, configuration

20   interaction (CI) analysis, and normal mode dynamics analysis. Computer programs for implementing these techniques are readily available. *See* Rein et al., *Computer-Assisted Modeling of Receptor-Ligand Interactions* (Alan Liss, New York, 1989). Compound derivatives are subjected to rescreening in the cell-based assay to select the one(s) that demonstrate the best interaction profile with the molecular target.
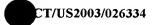
25

B. Compounds

The methods described herein are useful in screening a wide variety of compounds. For example, compounds to be screened in the present methods can be obtained from combinatorial libraries of peptides or small molecules, can be hormones, growth factors, and cytokines, can be

30   naturally occurring molecules or can be from existing repertoires of chemical compounds synthesized by the pharmaceutical industry. Combinatorial libraries can be produced for many types of compound that can be synthesized in a step-by-step fashion. Such compounds include polypeptides, beta-turn mimetics, polysaccharides, nucleic acids, phospholipids, hormones, prostaglandins, steroids, aromatic compounds, heterocyclic compounds, benzodiazepines,

35   oligomeric N-substituted glycines and oligocarbamates. Large combinatorial libraries of the compounds can be constructed by the encoded synthetic libraries (ESL) method described in

Affymax, WO 95/12608, Affymax, WO 93/06121, Columbia University, WO 94/08051, Pharmacopeia, WO 95/35503 and Scripps, WO 95/30642 (each of which is incorporated by reference for all purposes). Peptide libraries can also be generated by phage display methods. See, e.g., Devlin, WO 91/18980. Compounds to be screened can also be obtained from the National

5   Cancer Institute's Natural Product Repository, Bethesda, MD. Existing compounds or drugs with known efficacy can also be screened to evaluate side effects.

In addition to, or instead of, assessing mRNA or protein expression, a variety of different cellular and/or biochemical responses (also termed cell properties) can also be measured and compared in the methods described herein. For example, the cellular response to administration of

10   a compound can be quantified as a value or level of a cellular property, such as cell growth, neovascularization, hormone release, pH changes, changes in intracellular second messengers such as GMP, receptor binding and the like. The units of the value depend on the property. For example, the units can be units of absorbance, photon count, radioactive particle count or optical density.

15

Delivery

When the molecular target is intracellular, a compound that interacts with it must traverse the cell membrane. A compound contacted with a cell can cross the cell membrane in a number of ways. If the compound has suitable size and charge properties, it can be passively transported

20   across the membrane. Other processes of membrane passage include active transport (e.g., receptor mediated transport), endocytosis and pinocytosis. Where a compound cannot be effectively transported by any of the preceding methods, microinjection, biolistics or other methods can be used to deliver it to the internal portion of the cell. Alternatively, if the compound to be screened is a protein, a nucleic acid encoding the protein can be introduced into the cell and

25   expressed within the cell.

Likewise, the zinc finger protein-CME fusion to be tested must be introduced into the cell. Typically such is achieved by introducing either the ZFP-CME molecule or a nucleic acid encoding the ZFP-CME into the cell resulting in expression of the fusion protein within the cell. Nucleic acids can be introduced by conventional means including viral based methods, chemical methods,

30   lipofection and microinjection. The introduced nucleic acid can integrate into the host chromosome, persist in episomal form or can have a transient existence in the cytoplasm. Similarly, an exogenous protein can be introduced into a cell in protein form. For example, the zinc finger protein can be introduced by lipofection, biolistics, or microinjection or through fusion to membrane translocating domains.
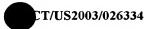
Thus, the compositions described herein can be provided to the target cell *in vitro* or *in vivo*. In addition, the compositions can be provided as polypeptides, polynucleotides or combination thereof.

5   A. Delivery of Polynucleotides

In certain embodiments, the compositions are provided as one or more polynucleotides. Further, as noted above, a zinc finger protein-containing composition can be designed as a fusion between a polypeptide zinc finger and one or more functional domains (*e.g.*, CMEs, activation domains and/or repression domains), that is encoded by a fusion nucleic acid. In both fusion and

10   non-fusion cases, the nucleic acid can be cloned into intermediate vectors for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors for storage or manipulation of the nucleic acid or production of protein can be prokaryotic vectors, (*e.g.*, plasmids), shuttle vectors, insect vectors, or viral vectors for example. A nucleic acid encoding a zinc finger protein can also cloned into an expression vector, for administration to a bacterial cell,

15   fungal cell, protozoal cell, piscine cell, plant cell, piscine cell, or animal cell, preferably a mammalian cell, more preferably a human cell.

To obtain expression of a cloned nucleic acid, it is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, *e.g.*, in Sambrook *et al.*, *supra*; Ausubel *et al.*, *supra*; and

20   Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990). Bacterial expression systems are available in, *e.g.*, *E. coli*, *Bacillus sp.*, and *Salmonella*. Palva *et al.* (1983) *Gene* 22:229-235. Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available, for example, from Invitrogen, Carlsbad, CA and Clontech, Palo Alto, CA.

25   The promoter used to direct expression of the nucleic acid of choice depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification. In contrast, when a protein is to be used *in vivo*, either a constitutive or an inducible promoter is used, depending on the particular use of the protein. In addition, a weak promoter can be used, such as HSV TK or a promoter having similar activity. The promoter

30   typically can also include elements that are responsive to transactivation, *e.g.*, hypoxia response elements, Gal4 response elements, *lac* repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system. *See, e.g.*, Gossen *et al.* (1992) *Proc. Natl. Acad. Sci USA* 89:5547-5551; Oligino *et al.*(1998) *Gene Ther.* 5:491-496; Wang *et al.* (1997) *Gene Ther.* 4:432-441; Neering *et al.* (1996) *Blood* 88:1147-1155; and Rendahl *et al.*

35   (1998) *Nat. Biotechnol.* 16:757-761.

In addition to a promoter, an expression vector typically contains a transcription unit or
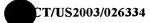
expression cassette that contains additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, *e.g.*, to the nucleic acid sequence, and signals required, *e.g.*, for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding, and/or translation
5   termination. Additional elements of the cassette may include, *e.g.*, enhancers, and heterologous spliced intronic signals.

A variety of inducible promoters (*e.g.*, operably linked to control expression of the ZFP-CME) can be used, for example using the tet-repressor system described in Gossen et al. Science (1995) 268:1766-1769, describe fusion of a tetracycline resistance gene repressor to a viral
10   transcription activation domain in order to induce rapid, greatly amplified gene expression in the presence of tetracycline. It is a modification of a preexisting system in which low levels of tetracycline prevented gene expression. The gene that codes for the tetracycline resistance gene repressor was mutagenized and a mutant fusion protein was created that depended on tetracycline for activation was identified. The construct can provide an on/off switch for high expression of a
15   gene.

Other activator/promoter sequences known in the art may also be used in construction of transactivator plasmids and plasmids in accordance with the present invention. These include, but are not limited to: (1) the T7 lac promoter construct activated by T7 RNA polymerase as the transactivator (Dubendorfs & Studier, J. Mol. Biol., 219: 45-49, 1991); (2) the Lex A (binding
20   domain)/Gal4 transcriptional activator-for the Lex A promoter (Brent & Ptashne, Cell 43: 729-736, 1985); (3) Gal4/VP16 (Carey et al., J- Mol. Biol. 209: 423-432, 1989; Cress et al., Science, 251: 87-90, 1991; Sadowski et al. Nature, 335: 563-564, 1988); (4) lac operator/repressor system as modified for eukaryotic expression (Brown et al., Cell 49: 603-612, 1987); (5) T7 polymerase-vaccinia virus promoter system (Fuerst et al., Proc. Natl. Acad. Sci. USA 83:
25   8122-8126; Fuerst et al., Molec. Cell Biol. 7: 2538-2544, 1987); (6) the T3 lac constructs activated by T3 RNA polymerase as the transactivator (Deuschle et al., Proc. Natl. Acad. Sci. USA 86: 5400-5404, 1989); and (7) glucocorticoid inducible mouse mammary tumor virus promoter system, (Lee et al., Nature 294: 228-232, 1981; Huang et al., Cell 27: 245-256, 1981; Ostrowski et al., Mol Cell. Biol. 3: 2045-2057, 1983). The tet operator/eCMV promoter exemplified herein also may be
30   modified to comprise the vaccinia virus promoter (Fuerst et al., 1987, supra) instead of the eCMV promoter.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the resulting ZFP polypeptide, *e.g.*, expression in plants, animals, bacteria, fungi, protozoa *etc.* Standard bacterial expression vectors include plasmids such
35   as pBR322, pBR322-based plasmids, pSKF, pET23D, and commercially available fusion expression systems such as GST and LacZ. Epitope tags can also be added to recombinant proteins

to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, *e.g.*, c-myc or FLAG.
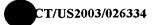
Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, *e.g.*, SV40 vectors, papilloma virus vectors, and vectors derived

5   from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

10   Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High-yield expression systems are also suitable, such as baculovirus vectors in insect cells, with a nucleic acid sequence coding for a ZFP as described herein under the transcriptional control of the polyhedrin promoter or any other strong baculovirus promoter.

15   Elements that are typically included in expression vectors also include a replicon that functions in *E. coli* (or in the prokaryotic host, if other than *E. coli*), a selective marker, *e.g.*, a gene encoding antibiotic resistance, to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the vector to allow insertion of recombinant sequences.

20   Standard transfection methods can be used to produce bacterial, mammalian, yeast, insect, or other cell lines that express large quantities of zinc finger proteins, which can be purified, if desired, using standard techniques. *See, e.g.,* Colley *et al.* (1989) *J. Biol. Chem.* **264:**17619-17622; and *Guide to Protein Purification, in Methods in Enzymology*, vol. 182 (Deutscher, ed.) 1990. Transformation of eukaryotic and prokaryotic cells are performed according to standard

25   techniques. *See, e.g.,* Morrison (1977) *J. Bacteriol.* **132:**349-351; Clark-Curtiss *et al.* (1983) in *Methods in Enzymology* **101:**347-362 (Wu *et al.*, eds).

Any procedure for introducing foreign nucleotide sequences into host cells can be used. These include, but are not limited to, the use of calcium phosphate transfection, DEAE-dextran-mediated transfection, polybrene, protoplast fusion, electroporation, lipid-mediated delivery (*e.g.*,

30   liposomes), microinjection, particle bombardment, introduction of naked DNA, plasmid vectors, viral vectors (both episomal and integrative) and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, *e.g.*, Sambrook *et al.*, *supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host

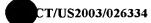35   cell capable of expressing the protein of choice.

Conventional viral and non-viral based nucleic acid delivery methods can be used to introduce nucleic acids into host cells or target tissues. Such methods can be used to administer nucleic acids encoding reprogramming polypeptides to cells *in vitro*. Additionally, nucleic acids are administered for *in vivo* or *ex vivo*. Non-viral vector delivery systems include DNA plasmids,

5   naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For reviews of nucleic acid delivery procedures, see, for example, Anderson (1992) *Science* 256:808-813; Nabel *et al.* (1993) *Trends Biotechnol.* 11:211-217; Mitani *et al.* (1993) *Trends Biotechnol.* 11:162-166; Dillon (1993) *Trends Biotechnol.*

10   11:167-175; Miller (1992) *Nature* 357:455-460; Van Brunt (1988) *Biotechnology* 6(10):1149-1154; Vigne (1995) *Restorative Neurology and Neuroscience* 8:35-36; Kremer *et al.* (1995) *British Medical Bulletin* 51(1):31-44; Haddada *et al.*, in *Current Topics in Microbiology and Immunology*, Doerfler and Böhm (eds), 1995; and Yu *et al.* (1994) *Gene Therapy* 1:13-26.

Methods of non-viral delivery of nucleic acids include lipofection, microinjection,

15   ballistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in, *e.g.*, U.S. Patent Nos. 5,049,386; 4,946,787; and 4,897,355 and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO

20   91/17424 and WO 91/16024. Nucleic acid can be delivered to cells (*ex vivo* administration) or to target tissues (*in vivo* administration).

The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to those of skill in the art. *See, e.g.*, Crystal (1995) *Science* 270:404-410; Blaese *et al.* (1995) *Cancer Gene Ther.* 2:291-297; Behr *et al.* (1994) *Bioconjugate*

25   *Chem.* 5:382-389; Remy *et al.* (1994) *Bioconjugate Chem.* 5:647-654; Gao *et al.* (1995) *Gene Therapy* 2:710-722; Ahmad *et al.* (1992) *Cancer Res.* 52:4817-4820; and U.S. Patent Nos. 4,186,183; 4,217,344; 4,235,871; 4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028 and 4,946,787.

The use of RNA or DNA virus-based systems for the delivery of nucleic acids take

30   advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to subjects (*in vivo*) or they can be used to treat cells *in vitro*, wherein the modified cells are administered to subjects (*ex vivo*). Conventional viral based systems for the delivery of ZFPs include retroviral, lentiviral, poxviral, adenoviral, adeno-associated viral, vesicular stomatitis viral and herpes viral

35   vectors. Integration in the host genome is possible with certain viral vectors, including the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long

term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, allowing alteration and/or expansion of the potential target cell population. Lentiviral vectors are

5 retroviral vector that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral nucleic acid delivery system would therefore depend on the target cell and/or tissue. Retroviral vectors have a packaging capacity of up to 6-10 kb of foreign sequence and are comprised of *cis*-acting long terminal repeats (LTRs). The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate

10 the exogenous gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), simian immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof. Buchscher *et al.* (1992) *J. Virol.* **66**:2731-2739; Johann *et al.* (1992) *J. Virol.* **66**:1635-1640; Sommerfelt *et al.* (1990) *Virol.* **176**:58-59; Wilson *et al.* (1989) *J. Virol.*

15 **63**:2374-2378; Miller *et al.* (1991) *J. Virol.* **65**:2220-2224; and PCT/US94/05700).

Adeno-associated virus (AAV) vectors are also used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* applications. *See, e.g.*, West *et al.* (1987) *Virology* **160**:38-47; U.S. Patent No. 4,797,368; WO 93/24641; Kotin (1994) *Hum. Gene Ther.* **5**:793-801; and Muzyczka (1994) *J. Clin. Invest.*

20 **94**:1351. Construction of recombinant AAV vectors are described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin *et al.* (1985) *Mol. Cell. Biol.* **5**:3251-3260; Tratschin, *et al.* (1984) *Mol. Cell. Biol.* **4**:2072-2081; Hermonat *et al.* (1984) *Proc. Natl. Acad. Sci. USA* **81**:6466-6470; and Samulski *et al.* (1989) *J. Virol.* **63**:3822-3828.

Recombinant adeno-associated virus vectors based on the defective and nonpathogenic

25 parvovirus adeno-associated virus type 2 (AAV-2) are a promising nucleic acid delivery system. Exemplary AAV vectors are derived from a plasmid containing the AAV 145 bp inverted terminal repeats flanking a transgene expression cassette. Efficient transfer of nucleic acids and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. Wagner *et al.* (1998) *Lancet* **351** (9117):1702-3; and Kearns *et al.* (1996)

30 *Gene Ther.* **9**:748-55. pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials. Dunbar *et al.* (1995) *Blood* **85**:3048-305; Kohn *et al.* (1995) *Nature Med.* **1**:1017-102; Malech *et al.* (1997) *Proc. Natl. Acad. Sci. USA* **94**:12133-12138. PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese *et al.* (1995) *Science* **270**:475-480. Transduction efficiencies of 50% or greater have been observed for MFG-S packaged

35 vectors. Ellem *et al.* (1997) *Immunol Immunother.* **44**(1):10-20; Dranoff *et al.* (1997) *Hum. Gene Ther.* **1**:111-2.

In applications for which transient expression is preferred, adenoviral-based systems are useful. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and are capable of infecting, and hence delivering nucleic acid to, both dividing and non-dividing cells. With such vectors, high titers and levels of expression have been obtained.

5 Adenovirus vectors can be produced in large quantities in a relatively simple system.

Replication-deficient recombinant adenovirus (Ad) vectors can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and/or E3 genes; the replication defector vector is propagated in human 293 cells that supply the required E1 functions in *trans*. Ad vectors can

10 transduce multiple types of tissues *in vivo*, including non-dividing, differentiated cells such as those found in the liver, kidney and muscle. Conventional Ad vectors have a large carrying capacity for inserted DNA. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection. Sterman *et al.* (1998) *Hum. Gene Ther.* 7:1083-1089. Additional examples of the use of adenovirus vectors for

15 nucleic acid delivery include Rosenecker *et al.* (1996) *Infection* 24:5-10; Sterman *et al.*, *supra*; Welsh *et al.* (1995) *Hum. Gene Ther.* 2:205-218; Alvarez *et al.* (1997) *Hum. Gene Ther.* 5:597-613; and Topf *et al.* (1998) *Gene Ther.* 5:507-513.

Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and Ψ2 cells or PA317 cells, which

20 package retroviruses. Viral vectors used in nucleic acid delivery are usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. Missing viral functions are supplied in *trans*, if necessary, by the packaging cell line. For example,

25 AAV vectors used in nucleic acid delivery typically only possess ITR sequences from the AAV genome, which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the

30 helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment, which preferentially inactivates adenoviruses.

In many nucleic acid delivery applications, it is desirable that the vector be delivered with a high degree of specificity to a particular tissue type. A viral vector can be modified to have

35 specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the outer surface of the virus. The ligand is chosen to have affinity for a receptor known to be
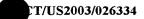
present on the cell type of interest. For example, Han *et al.* (1995) *Proc. Natl. Acad. Sci. USA*
92:9747-9751 reported that Moloney murine leukemia virus can be modified to express human
heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells
expressing human epidermal growth factor receptor. This principle can be extended to other pairs

5   of virus expressing a ligand fusion protein and target cell expressing a receptor. For example,
filamentous phage can be engineered to display antibody fragments (*e.g.*, F_{ab} or F_v) having specific
binding affinity for virtually any chosen cellular receptor. Although the above description applies
primarily to viral vectors, the same principles can be applied to non-viral vectors. Such vectors can
be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

10   Vectors can be delivered *in vivo* by administration to a subject, typically by systemic
administration (*e.g.*, intravenous, intraperitoneal, intramuscular, subdermal, or intracranial
infusion) or topical application, as described *infra*. Alternatively, vectors can be delivered to cells
ex vivo, such as cells explanted from a subject (*e.g.*, lymphocytes, bone marrow aspirates, tissue
biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a

15   subject, usually after selection for cells which have incorporated the vector.

         Ex vivo cell transfection (*e.g.*, for diagnostics, research, or for gene therapy such as via re-
infusion of the transfected cells into the host organism) is well known to those of skill in the art. In
a preferred embodiment, cells are isolated from the subject organism, transfected with a nucleic
acid (gene or cDNA), and re-infused back into the subject organism (*e.g.*, patient). Various cell

20   types suitable for *ex vivo* transfection are well known to those of skill in the art. *See, e.g.*, Freshney
et al., *Culture of Animal Cells, A Manual of Basic Technique,* 3rd ed., 1994, and references cited
therein, for a discussion of isolation and culture of cells from patients.

         In one embodiment, hematopoietic stem cells are used in *ex vivo* procedures for cell
transfection and nucleic acid delivery. The advantage to using stem cells is that they can be

25   differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor
of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ stem
cells *in vitro* into clinically important immune cell types using cytokines such a GM-CSF, IFN-γ
and TNF-α are known. Inaba *et al.* (1992) *J. Exp. Med.* 176:1693-1702.

         Stem cells are isolated for transduction and differentiation using known methods. For

30   example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with
antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells),
GR-1 (granulocytes), and Iad (differentiated antigen presenting cells). *See* Inaba *et al., supra.*

         Vectors (e.g., retroviruses, adenoviruses, liposomes, *etc.*) containing nucleic acids can be
also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked

35   DNA can be administered. Administration is by any of the routes normally used for introducing a
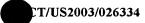molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such

nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition

5   being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions described herein. *See, e.g., Remington's Pharmaceutical Sciences,* 17th ed., 1989.

B. Delivery of Polypeptides

10   In other embodiments, fusion proteins are administered directly to target cells. In certain *in vitro* situations, the target cells are cultured in a medium containing one or more CMEs (or functional fragments thereof) fused to one or more of the ZFPs described herein. In other situations, fusion proteins can be administered to cells or tissues *in vivo* or *ex vivo*.

An important factor in the administration of polypeptide compounds is ensuring that the

15   polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins, lipids and other compounds, which have the ability to translocate polypeptides

20   across a cell membrane, have been described.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, *Antennapedia,* was found to be the third

25   helix of the protein, from amino acid position 43 to 58. Prochiantz (1996) *Curr. Opin. Neurobiol.* 6:629-634. Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics. Lin *et al.* (1995) *J. Biol. Chem.* **270**:14255-14258.

Examples of peptide sequences which can be linked to a zinc finger polypeptide (or fusion

30   containing the same) for facilitating its uptake into cells include, but are not limited to: an 11 amino acid peptide of the *tat* protein of HIV; a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (*see* Fahraeus *et al.* (1996) *Curr. Biol.* 6:84); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi *et al.* (1994) *J. Biol. Chem.* 269:10444); the h region of a signal peptide, such as the Kaposi fibroblast growth factor (K-FGF) h

35   region (Lin *et al., supra*); and the VP22 translocation domain from HSV (Elliot *et al.* (1997) *Cell*
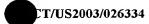
88:223-233). Other suitable chemical moieties that provide enhanced cellular uptake can also be linked, either covalently or non-covalently, to the ZFPs.

Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules (called "binary toxins") are composed of at least two parts: a translocation
5   or binding domain and a separate toxin domain. Typically, the translocation domain, which can optionally be a polypeptide, binds to a cellular receptor, facilitating transport of the toxin into the cell. Several bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), have been used to deliver peptides to the cell cytosol as internal or
10   amino-terminal fusions. Arora *et al.* (1993) *J. Biol. Chem.* 268:3334-3341; Perelle *et al.* (1993) *Infect. Immun.* 61:5147-5156; Stenmark *et al.* (1991) *J. Cell Biol.* 113:1025-1032; Donnelly *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:3530-3534; Carbonetti *et al.* (1995) *Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295; Sebo *et al.* (1995) *Infect. Immun.* 63:3851-3857; Klimpel *et al.* (1992) *Proc. Natl. Acad. Sci. USA.* 89:10277-10281; and Novak *et al.* (1992) *J. Biol. Chem.* 267:17186-
15   17193.

Such subsequences can be used to translocate polypeptides, including the polypeptides as disclosed herein, across a cell membrane. This is accomplished, for example, by derivatizing the fusion polypeptide with one of these translocation sequences, or by forming an additional fusion of the translocation sequence with the fusion polypeptide. Optionally, a linker can be used to link the
20   fusion polypeptide and the translocation sequence. Any suitable linker can be used, *e.g.*, a peptide linker.

A suitable polypeptide can also be introduced into an animal cell, preferably a mammalian cell, via liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate
25   an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell.

The liposome fuses with the plasma membrane, thereby releasing the compound into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome is either degraded or it fuses with the membrane of the transport vesicle and releases its contents.
30   In current methods of drug delivery *via* liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer is degraded over time through the action of various agents in the body. Alternatively, active drug release involves using an agent to induce a permeability change in the
35   liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near the liposome membrane. *See, e.g., Proc. Natl. Acad. Sci.*

USA 84:7851 (1987); *Biochemistry* 28:908 (1989). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

5   For use with the methods and compositions disclosed herein, liposomes typically comprise a fusion polypeptide as disclosed herein, a lipid component, *e.g.*, a neutral and/or cationic lipid, and optionally include a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (*e.g.*, an antigen). A variety of methods are available for preparing liposomes as described in, *e.g.*; U.S. Patent Nos. 4,186,183; 4,217,344; 4,235,871;

10   4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028; 4,235,871; 4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028; 4,946,787; PCT Publication No. WO 91/17424; Szoka *et al.* (1980) *Ann. Rev. Biophys. Bioeng.* 9:467; Deamer *et al.* (1976) *Biochim. Biophys. Acta* 443:629-634; Fraley, *et al.* (1979) *Proc. Natl. Acad. Sci. USA* 76:3348-3352; Hope *et al.* (1985) *Biochim. Biophys. Acta* 812:55-65; Mayer *et al.* (1986) *Biochim. Biophys. Acta* 858:161-168; Williams *et*

15   *al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:242-246; *Liposomes*, Ostro (ed.), 1983, Chapter 1); Hope *et al.* (1986) *Chem. Phys. Lip.* 40:89; Gregoriadis, *Liposome Technology* (1984) and Lasic, *Liposomes: from Physics to Applications* (1993). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are well known in

20   the art.

        In certain embodiments, it may be desirable to target a liposome using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (e.g., ligands, receptors, and monoclonal antibodies) has been previously described. *See, e.g.*, U.S. Patent Nos. 4,957,773 and 4,603,044.

25   Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as *ras* or *c-erbB2*. In addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and carcinoembryonic antigen (CEA). Sites of

30   viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV-1) and papilloma virus antigens. Inflammation can be detected using molecules specifically recognized by surface molecules which are expressed at sites of inflammation such as integrins (e.g., VCAM-1), selectin receptors (e.g., ELAM-1) and the like.

35   Standard methods for coupling targeting agents to liposomes are used. These methods generally involve the incorporation into liposomes of lipid components, *e.g.*,

phosphatidylethanolamine, which can be activated for attachment of targeting agents, or incorporation of derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A. *See* Renneisen *et al.* (1990) *J. Biol. Chem.* 265:16337-16342 and Leonetti *et al.* (1990) *Proc. Natl.*

5 *Acad. Sci. USA* 87:2448-2451.

Kits

Also provided are kits for performing any of the above methods. The kits typically contains cells comprising a ZFP-CME fusion polypeptide and/or a nucleic acid encoding a ZFP-

10 CME fusion polypeptide for use in the above methods, or components for making such cells. Some kits contain pairs of test and control cells differing in that one cell population is transformed with an exogenous nucleic acid encoding a ZFP-CME fusion protein designed to regulate expression of a molecular target or other protein within the test cells. Some kits contain a single cell type and other components that allow one to produce control cells from that cell type. Such

15 components can include a vector encoding a zinc finger protein or the zinc finger protein itself. Additional kits contain nucleic acids which encode one or more ZFP-CME fusion proteins. The kits can also contain buffers for transformation of cells, culture media for cells, and/or buffers for performing assays. Typically, the kits also contain a label indicating that the cells are to be used for screening compounds. A label includes any material such as instructions, packaging or

20 advertising leaflet that is attached to or otherwise accompanies the other components of the kit.
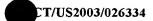

Additional Applications

In addition to identifying novel modulators (*e.g.,* activators, inhibitors) of various chromatin-modifying activities, the compositions and methods described herein allow for the

25 identification of whether specific modifications act as dominant signals to regulate gene expression *in vivo*, which, in turn, functionally validates the importance of specific enzymes in gene regulation.

Furthermore, the identification of compounds that modulate chromatin modifying enzymatic activities can be therapeutically applied, for example using CME activators to stimulate

30 the differentiation of a specific cell lineage, as an anti-proliferation/cancer therapy or in the treatment of specific diseases. Acetylation of the tat protein of HIV (via p/CAF, GCN5), for example, is required for functional tat and the identification of compounds that inhibit tat acetylation allows for the development of drugs that prevent and/or treat HIV infection.

Similarly, the development and progression of a significant proportion of human cancers,

35 especially those of the lymphoid lineage, are thought to be mediated by alterations in the normal function and targeting of chromatin modifying activities. Examples include the chromosomal

translocations and fusions of the AML and PML oncoproteins. The identification of compounds
that modulate CMEs as described herein (*e.g.*, by substitution of the DNA-binding domain of an
oncogenic fusion transcription factor with an engineered binding domain targeted to a reporter
gene, followed by an assay for modulators of reporter gene expression) will find use in the

5   development of drug therapies for leukemias and other cancers. *See*, for example, Cairns (2001)
    Trends Cell Biol. **11**:S15-S21; Jones *et al.* (2002) *Br. J. Haematol.* **118**:714-727; Bain (2002) *Acta
    Haematol.* **107**:57-63; Lorsbach *et al.* (2001) *Intl. J. Haematol.* **74**:258-2665; Lin *et al.* (2001)
    Oncogene **20**:7204-7215; Alcalay *et al.* (2001) *Oncogene* **20**:5680-5694; Licht (2001) *Oncogene*
    20:5660-5679; Hiebert *et al.* (2001) *Curr. Opin. Haematol.* **8**:197-200; Crans *et al.* (2001)

10 *Leukemia* **15**:313-331; Lutterbach *et al.* (2000) *Gene* **245**:223-235; and Look (1997) *Science*
    278:1059-1064.

        In addition, the disclosed fusion molecules can be used *in vitro* to test for the effect of a
    compound on the catalytic activity of the CME portion of the fusion. See Example 2.

15   EXAMPLES
        Introduction
            Histone H3 and H4 are methylated at lysine and arginine residues (7). These modifications
    have been associated with both transcriptionally active and repressed genes *in vivo*. Methylation of
    specific arginine residues by the coactivator-associated proteins CARM1 and PRMT1 has been

20 associated with transcriptional activation (8-10). Methylation of lysine 4 of histone H3, however,
    has been implicated in both transcriptional activation (11, 12) and in telomeric silencing (13, 14),
    but primarily the methylation of residues within histone H3 has been linked to transcriptional
    repression. One such modification, the methylation of lysine 9 of histone 3 (H3K9) has received
    particular attention following the discovery of multiple enzymes that catalyze this modification.

25 The first such methyltransferases discovered, SUV39H1 and SUV39H2 (15, 16), are close
    homologues of the *Drosophila* heterochromatin associated protein Su(var)3-9, a modifier of
    position effect variegation (17, 18), and the *S. pombe* silencing factor Clr4 (19) and thus formally
    connect HMTs with the regulation of chromatin structure (20). Furthermore, several labs have
    shown histone H3 methylated on K9 to be specifically bound by the chromodomain of HP1 (21,

30 22), a protein implicated both in heterochromatic silencing and gene repression. These data
    suggested that methylation of histone H3K9 serves mechanistically as an epigenetic signal,
    catalyzing the recruitment of heterochromatin proteins leading to transcriptional repression. In
    support of this model the repression of the E2F regulated cyclin E promoter via the Retinoblastoma
    protein (Rb) depends in part upon its recruitment of the histone methyltransferase SUV39H1 and

35 associated HP1 (23). Additionally, the recent purification of an E2F6 associated complex in
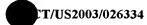   quiescent cells (24) that contains both the potent H3K9 methyltransferase G9A (25) and its close

homologue Eu-HMTase1 further implicates H3K9 HMTs in the regulation of E2F target genes. Importantly, transcriptional control mediated through the KRAB repression domain, one of the most common repression domains found in DNA binding proteins from mammalian cells, has been recently linked to the recruitment of a different H3K9 methyltransferase - SETDB1 (26). Thus, the

5 methyl-H3K9 epigenetic imprint is not limited to large heterochromatic domains but appears to play a general role in the regulation of gene expression.

While these reports support the correlation between specific histone methylation patterns and transcriptional repression, the question remains as to whether histone methylation is causative in the initiation and establishment of gene repression, or is instead a byproduct of the process

10 leading to the repressed state. In order to better understand the process of transcriptional repression and to address directly the role of histone methylation in this process, we have employed our ability to design synthetic zinc-finger transcription factors with novel binding specificities. We have previously applied this capability to achieve the regulation of endogenous genes both in the activation of silent genes e.g. EPO1 (27), the activation of inducible loci e.g. VEGF-A (28) and in

15 the repression of gene expression e.g. PPARγ (29). Using ZFPs which regulate the expression of the endogenous VEGF-A gene (28), we show here that direct targeting of the minimal catalytic HMT domain of either SUV39H1 or G9A alone is sufficient to effect the local methylation of H3K9 at the promoter of this locus and the consequent transcriptional repression of the VEGF-A gene. In keeping with the *in vitro* substrate specificity of these HMTs, we observe that ZFP-HMT

20 mediated repression is enhanced when coupled with histone deacetylase recruitment, supporting genetic studies which connect H3K9 HMT activity with histone deacetylation. Importantly, amino acid substitutions within the catalytic core of SUV39H1 (15), which abolish the methylation activity of this domain, eliminate the ability of these chimeric transcription factors to repress transcription and abolish VEGF-A promoter histone methylation. Taken together, these data

25 suggest that histone methylation is a primary signal sufficient to drive a targeted repression pathway capable of inhibiting the transcription of endogenous genes *in vivo*.

Example 1: Targeted Histone Methyltransferase (HMT) Domains Repress Gene Expression
in vivo

30 To determine whether methylation of the histone tails is causative in the process of gene repression, we linked the catalytic HMT activities of both SUV39H1 and G9A to an engineered zinc-finger transcription factor (VZ+434 [28] referred to here as ZFP-A). This engineered DNA binding protein specifically recognizes a site at position +434 relative to the transcription start site of the endogenous human VEGF-A locus, and has been shown previously to up-regulate VEGF-A

35 transcription when fused to the VP16 and p65 activation domains. This approach therefore allows us to use an endogenous chromosomal gene as a transcriptional reporter system. All three
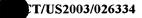
constructs, G9A and both the SUV39H1 deletion 76 and deletion 149 (referred to throughout as

Suv Del 76 or Suv Del 149) constructs employed in this study contain the minimal catalytically

active portions of the proteins [15, 25] (shown schematically in Fig. 1A). Specifically, the ZFPA-

SUV39H1 constructs used lack the chromodomain, the HP1 interaction region and the amino-

5   terminal transcriptional repression domain of SUV39H1. Fig.1B and C show that chimeras of

ZFP-A with either G9A, or the two SUV39H1 deletions were all able to efficiently repress the

amount of VEGF-A protein (Fig.1B) and mRNA (Fig.1C) produced by the endogenous VEGF-A

locus ~2 and ~3 fold respectively, despite background VEGF-A gene expression from non-

transfected cells. Fusion of an alternative repression domain encoding the LBD of v-ErbA, a viral

10   relative of avian thyroid hormone receptor protein and a known HDAC3/NCoR recruitment

domain, to ZFPA led to a similar decrease in transcription from this locus (Fig. 1B and C).

Furthermore, use of a ZFP that targets these domains to the IGF2 and H19 genes led to the

concomitant transcriptional repression of these specific targets, demonstrating the generality of this

effect. No repression was observed when HEK293 cells were transfected with a plasmid

15   expressing v-ErbA LBD fused to GFP, nor with cells transfected with control plasmids

(PCDNA3.1 and GFP). All ZFPs were expressed to similar degrees in this cell line (Fig.1E). Thus

direct recruitment of HMT domains results in the transcriptional repression of the targeted gene *in*

vivo. Interestingly, however, neither G9A, SUV39H1 nor v-ErbA ZFP fusion proteins were able to

effectively repress a transiently transfected luciferase reporter plasmid containing the VEGF-A

20   gene promoter (Fig. 1D and [30]), although significant upregulation of luciferase activity was

observed when equivalent amounts of ZFP-A fused to the p65 activation domain was used (Fig.

1D). This result supports the chromatin-specific mode of repression expected for the ZFP-HMTs,

and highlights the utility of assays employing endogenous genes.

25   **Example 2: HMT Activity is Necessary for Gene Repression**

            To provide evidence for the direct role of histone methyltransferase activity in the

repression observed at the endogenous VEGF-A locus, we first confirmed that the ZFP-domain

chimeras were catalytically active *in vitro*. HA epitope tagged ZFP-HMTs were

immunoprecipitated from transfected cell extracts and assayed for HMT activity. Fig.2A (Panel I)

30   demonstrates that both the Suv Del 76 and G9A ZFP chimeras have intrinsic HMT activity, while

control precipitations and those utilizing extracts transfected with the ZFP DNA binding domain

alone did not precipitate HMTase activity. As has been previously reported, G9A is a significantly

more potent HMT than SUV39H1 in this assay. A western blot of the immunoprecipitates

employed in the *in vitro* HMTase assay shown in Fig.2A (Panel II) confirms that this difference in

35   activity between the G9A and Suv Del 76 chimeras is not a consequence of differential expression

levels.

We next asked whether the HMT activity the ZFP-fusions possess *in vitro* was necessary
for the repression observed at the endogenous VEGF-A promoter *in vivo*. To address this issue,
point mutations were constructed which abolish the HMT activity of the domains [15]. Three
different amino acid substitutions were created in the context of the ZFPA-SUV Del 76 fusion:

5   H324K (mutant A), C326A (mutant B), and both mutations in combination (mutant AB), shown in
Fig. 2B. When assayed by transient transfection, mutations in either site alone or the combination
largely abolished the transcriptional repression function of the ZFP-chimeras at the endogenous
VEGF-A gene *in vivo* (Fig. 2C). This loss of repression function was not a consequence of poor
expression of the mutant chimeras relative to wild type (Fig.2D), but rather is a direct consequence

10   of ablation of their HMT activity [15]. Interestingly, full length SUV39H1 fused to ZFPA was
unable to efficiently drive the repression of the VEGF-A promoter, in sharp contrast to the fusions
comprising only the catalytic domain (Fig. 2C). This is perhaps due to the presence of an intact
HP1 interaction domain [31, 32], which could recruit and hence mis-target the ZFP-SUV39H1
chimeras to regions of heterochromatin, thus competing with the ZFP DNA binding domain and

15   precluding the localization of the protein at the VEGF-A gene.

**Example 3: Increased Transcriptional Repression Through Direct Recruitment of HMTs and
HDAC Activities**

The precise modification state of the histone tails can influence potential subsequent

20   modifications. For example, both the acetylation and phosphorylation status of nearby residues
significantly adversely affects SUV39H1 HMT activity *in vitro* [15, 20], and mutation of the
HDAC Clr3 compromises global H3K9 methylation of heterochromatic domains in *S. pombe* [20].
We reasoned that if acetylation levels at the VEGF-A promoter were reduced via the direct
targeting of an HDAC activity, this might increase the efficacy of the HMT domains catalytic

25   function, and thus increase their repression potential. To test this model we employed a second
designed ZFP that also regulates the transcription of the VEGF-A gene (Fig. 3A, and [28]). ZFP-
A, which has been used throughout this study thus far, binds between a pair of binding sites for a
second engineered ZFP (VZ +42/ +530 referred to as ZFP-B). Importantly, these two proteins
have binding sites which are separated by less than a nucleosome length, thus ensuring that the

30   catalytic activities recruited to either location should act upon the same local region of chromatin
(Fig. 3A). To recruit HDAC activity to the promoter we employed the v-ErbA repression domain.
v-ErbA is a viral relative of avian thyroid hormone receptor (TR) that constitutively recruits the
NCoR/SMRT corepressor complex, which utilizes associated HDAC activities to repress
transcription (see [33]) and references therein). Repression of VEGF-A by v-ErbA (Fig. 1) is

35   associated with consequent deacetylation of the promoter proximal histone tails. Fig. 3B shows
that G9A, Suv Del 76 and vErbA function to repress the endogenous VEGF-A locus when fused to

either of the two ZFPs. GFP fusions of all three domains or ZFP alone constructs (without a functional domain) fail to affect the transcription of this locus. Notably, recruitment of both G9A and v-ErbA simultaneously to the VEGF-A promoter resulted in a significant increase in the level of repression, beyond that of either functional domain alone (Fig.3C and Experimental

5   Procedures). Importantly, no increase in repression was observed by the use of combinations of ZFPs fused to the same domain (Fig.3D). Thus repression by HMT domains is enhanced by the simultaneous delivery of histone deacetylase function, confirming the functional link between these chromatin modifications in gene repression *in vivo*.

10   **Example 4: Targeting of HMT domains to the Endogenous VEGF-A Promoter Results in H3 Lysine 9 Methylation**

We have shown that ZFP-HMT catalytic domains fusions derived from either SUV39H1 or G9A are able to repress the expression of the endogenous VEGF-A gene, in a manner that depends upon their catalytic activity (Fig. 1 & 2). To confirm that the repression observed resulted from the

15   methylation of histone H3 lysine 9 *in vivo*, we performed chromatin immunoprecipitation (ChIP) assays with antibodies specific for dimethyl-H3K9 (Upstate Biotechnology Inc). Precipitated DNA was then analyzed with primers specific for a region adjacent to the ZFP binding site (+400), and internally controlled using primers for exon 7 of the GAPDH locus. The results of this analysis are shown in Fig. 4A. The level of H3K9 methylation signal is enriched ~2-3 fold by ZFP-Suv Del 76

20   or ZFP-G9A compared to either the mock transfected or the ZFPA no functional domain control samples. No signal was obtained from control samples prepared in the absence of the antibody. Importantly, transfection with the ZFP linked to the catalytically null HMT mutant abolished the enrichment of dimethly-H3K9 leaving only a residual signal when compared to the control samples (Fig. 4A). Control primers for exon 1 of the p16 locus (non-methylated and expressed in HEK293

25   cells) show no enrichment upon introduction of the ZFP-HMT fusions again demonstrating that the enrichment is specific to the VEGF-A promoter (Fig. 4A). Thus, direct targeting of HMT activity to the VEGF-A promoter via a designed zinc-finger transcription factor results in the specific methylation of H3K9 within nucleosomes proximal to the ZFP binding site *in vivo*.

Methylation of H3K9 is postulated to act as a signal tag for the recruitment of HP1, which

30   through self-association and interaction with SUV39H1 is thought to result in the spread of the methylation signal away from the original methylated H3K9 residue/HP-1 binding site [21]. To determine if this spreading occurs when the ZFP-HMT chimeras are employed, we re-analyzed the above samples with primer sets specific for regions increasingly distant from the ZFP binding site, i.e. +1, and -500, relative to the transcription start site. The results, shown in Figs. 4B and 4C,

35   demonstrate that both ZFP-G9A and ZFP-SUV Del 76 are able to increase the H3K9 methylation signal at the VEGF-A promoter at all primer-probe sets used. This effect is not observed at VEGF-

A when constructs without a functional domain, without a ZFP DNA binding domain or without catalytic HMT function are employed (Figs. 4B and 4C). Taken together these experiments demonstrate the ability of ZFP-HMT fusions to generate a histone H3K9 methylation signal specifically at the promoter targeted by the ZFP, and that this methylation imprint is not restricted to the immediate vicinity of the ZFP binding site.

5

We show here how novel ZFP transcription factors can be engineered to direct functional domains and/or catalytic activities to precise locations within the human genome to regulate endogenous gene expression. In contrast to transient reporter gene assay, this unique capability allows the study of the molecular processes involved in both transcriptional regulation and in the modulation of chromatin structure directly on endogenous genes *i.e.* within their native chromatin architecture. Furthermore, the direct targeting of enzymatic activities such as the H3K9 HMTs described here, able to epigenetically mark a region of chromatin for transcriptional silencing provides both a powerful research tool and potential therapeutic avenue in the study and treatment of human disease.

10

15

Example 5: Experimental Procedures

Cell Culture and Transient Transfections

HEK293 cells were grown in Dulbecco's modified eagle medium supplemented with 10% fetal bovine serum in a 5% CO2 incubator at 37 C. For transfections, HEK293 cells were plated in 12-well plates at a density of 250,000 cells/well and transfected 1 day later using Lipofectamine 2000 reagent (Gibco-BRL, MD) according to manufacturers recommendations, using 9 μl of Lipofectamine 2000 reagent and 1.5 μg of ZFP plasmid DNA per well. The medium was removed and replaced with fresh medium 6-12 h after transfection. For the co-expression experiments in Fig. 3, to compensate for the difference in apparent Kd of each ZFP for its binding site [27], 250ng of the ZFP-A constructs were combined with 1.25μg of the ZFP-B constructs throughout.

20

25

Immunoprecipitations and histone methyltransferase activity assays

Whole cell lysates were pre-cleared using 40ul bed volume of Protein G Agarose beads for 30 minutes at 4°C with agitation. The lysates were spun at 1000rpm for 1 minute and the clarified lysate removed and used in an immunoprecipitation with 5ul/ of anti-HA epitope tag antibody (sc-7392 Santa Cruz) or an IgG control antibody and incubated at 4°C for 2hrs with agitation. 20ul bed volume of Protein G Agarose beads was then added and the samples incubated for an additional hour at 4 C with agitation. Samples were spun down (1000rpm for 1minute) and washed x3 in wash buffer (20mM HEPES pH7.9, 75mM KCl, 2.5mM MgCl₂, 1mM DTT, 0.5mM PMSF). Samples were then re-suspended in HMT assay buffer (50mM TRIS pH8.0, 20mM KCl, 250mM Sucrose, 10mM MgCl₂, 1mM DTT) containing 10μg bulk histones (Sigma) and 1ul S-adenosyl -

30

35

(methyl-^{3}H)-L-methionine (80 Ci/mM; PerkinElmer life sciences) as methyl donor. Samples were incubated for 1hr at 30 °C, and reactions terminated by addition of x2 SDS PAGE sample buffer. Protein samples were resolved on a 10-20% SDS PAGE gradient gel (Biorad) which was then treated with Amplify (Amersham Pharmacia Biotech), dried down and exposed to X-Ray film.

5

Design of ZFP DNA-binding domains

ZFP DNA-binding domains were designed and synthesized according to co-owned WO 00/41566, WO 00/42219 and WO 02/46412. See also Zhang *et al.* (2000) *J. Biol. Chem* **275**: 33,850-33,860 and Liu *et al.* (2001) *J. Biol. Chem* **276**:11323-11334.

10

rtPCR cloning of SUV39H1 and G9A.

cDNA was prepared and pooled from RNA derived from HEK293, MCF-7 and U2OS cell lines using the thermoscript RT-PCR system (GibcoBRL) according to manufacturers recommendations. The DNA encoding human G9A catalytic domain and full length SUV39H1

15 were generated via PCR from the cDNA pool using Platinum Taq High Fidelity DNA polymerase (Invitrogen) using the indicated primers. Constructs generated via PCR were sequence confirmed. SUV39H1 deletions 76 and 149 were generated via PCR using Full length SUV39H1 as template. All HMT constructs were subcloned into ZFP plasmid backbone using engineered 5' BamH1 and 3' XhoI restriction sites.

20 rt-PCR full length hSuv39H1 forward primer CGGATCCCCGTGGGGAAAGATGGCGG (SEQ ID NO:14), reverse primer CGCGGCCGCGACAGGAGGGCAGCAGTGGG (SEQ ID NO:15)
Suv39H1 Deletion 76 forward primer CAGGATCCCCACGGCAGAATCTCAA (SEQ ID NO:16)
Suv39H1 Deletion 149 forward primer CAGGATCCGAGAATGAGGTGGACCTG (SEQ ID NO:17)
Reverse Suv39H1 Deletion primer TACTCGAGCTAGAAGAGGTATTTGCGGCAGGACTC (SEQ ID NO:18)
25 G9A *cat* domain forward primer GAGGATCCGGCAGCGCCGCCATCGCCGAA (SEQ ID NO:19), reverse primer GACTCGAGTCATGTGTTGACAGGGGGCAGG (SEQ ID NO:20)

Site directed mutagenesis of Suv39H1 Del 76

Site directed mutagenesis was performed using the quick change site directed mutagenesis
30 kit (Stratagene) according to manufacturers recommendations.
mutant A forward primer CTCCCACTTTGTCAACAAAAGTTGTGACCCCAACCTGCAG (SEQ ID NO:21)
mutant A reverse primer CTGCAGGTTGGGGTCAACAACTTTTGTTGACAAAGTGGGAG (SEQ ID NO:22)
mutant B forward primer CTCCCACTTTGTCAACCACAGTGCTGACCCCAACCTGCAG (SEQ ID NO:23)
mutant B reverse primer CTGCAGGTTGGGGTCAGCACTGTGGTTGACAAAGTGGGAG (SEQ ID NO:24)
35 mutant AB forward primer CTCCCACTTTGTCAACAAAAGTGCTGACCCCAACCTGCAG (SEQ ID NO:25)
mutant AB reverse primer CTGCAGGTTGGGGTCAGCACTTTTGTTGACAAAGTGGGAG (SEQ ID NO:26)

Secreted VEGF-A protein ELISA

Secreted VEGF-A in the tissue culture media by transfected HEK293 cells was assayed
after 48 hrs using a human VEGF-A ELISA kit (R&D systems) in duplicate according to
5   manufacturers recommendations.

Immunoblot detection of ZFP fusion proteins

For western analysis of ZFP protein expression, cells were lysed 72 hrs post transfection in
RIPA buffer (Santa Cruz) according to recommendations. Samples were boiled in x2 Laemmli
10   sample buffer and resolved by SDS PAGE, followed by western blotting using an anti-HA epitope
tag antibody (sc-7392 Santa Cruz). The westerns were visualized by ECL (Amersham Pharmacia
Biotech) as described previously [35]

Quantitative RT-PCR analysis of VEGF-A mRNA expression (TaqMan)

15   HEK293 cells were lysed and total RNA prepared using the high pure RNA isolation kit
(Roche) according to manufacturers recommendations. RNA (25 ng) was used in real-time
quantitative RT-PCR analysis using TaqMan chemistry in a 96-well format on an ABI 7700 SDS
machine (PerkinElmer Life Sciences) as described previously [27]. Briefly, reverse transcription
was performed at 48 °C for 30 min using MultiScribe reverse transcriptase (PerkinElmer Life
20   Sciences). Following a 10-min denaturation at 95 °C, PCR amplification using AmpliGold DNA
polymerase was conducted for 40 cycles at 95 °C for 15 s and at 60 °C for 1 min. Primer/probes
used were as described previously [28]. The results were analyzed using SDS Version 1.6.3
software.

25   *Chromatin immunoprecipitation (ChIP)*

Chromatin immunoprecipitation was performed using the ChIP assay kit according to the
manufacturer's instructions (Upstate Biotechnology, NY). Briefly, approximately 10 million cells
were transfected with expression plasmid and utilized 72 hrs post transfection. Samples were
cross-linked with 1% formaldehyde for 10 min at 37°C, washed with PBS, and re-suspended in
30   lysis buffer. The cell lysate was sonicated on ice for a total of 2 minutes (in 5 second pulses),
resulting in an average DNA fragment length of approximately 500 bp. After removing cell debris
by centrifugation and pre-clearing of lysate, immunoprecipitation was performed in ChIP dilution
buffer overnight with anti-dimethyl-Histone H3K9 (Upstate Biotechnology) with agitation.
Salmon sperm DNA/Protein A agarose slurry was added and incubated for 1hr at 4°C with
35   agitation. The antibody–agarose complex was centrifuged and washed x5, and the
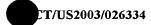immunoprecipitated fraction eluted. The cross-linking was reversed by incubation at 65°C for 4

hours in the presence of 200 mM NaCl. The DNA was recovered by phenol/chloroform extraction, precipitated and the abundance of specific sequences quantitated using real time PCR (TaqMan) as described above, omitting the reverse transcription reaction step. Relative abundances of the various VEGF-A genomic primers were calculated relative to an internal GAPDH genomic probe

5   set.

GAPDH genomic forward primer ACATCAAGAAGGTGGTGAAG (SEQ ID NO:27), reverse primer AGCTTGACAAAGTGGTCGTTG (SEQ ID NO:28), VEGF +400 region forward primer CAGCGAAAGCGACAGGGG (SEQ ID NO:29), reverse primer GTCAGCTGCGGGATCCC (SEQ ID NO:30), VEGF −500 region forward primer GGCCACCACAGGGAAGCT (SEQ ID NO:31), reverse primer

10   ACACAGACACACACGTCCTCACT (SEQ ID NO:32), VEGF start site (+1 region) forward primer AGGATCGCGGAGGCTTG (SEQ ID NO:33), reverse primer CGACAGAGCGCTGGTGCTA (SEQ ID NO:34), p16 exon 1 forward·primer TCTGGAGGACGAAGTTTGCA (SEQ ID NO:35), reverse primer CCAGGAAGCCTCCCCTTTT (SEQ ID NO:36).

15   **Example 6: Regulation of endogenous reporter genes by inducible expression of ZFP-HMT fusions**

Clonally isolated TREx U2OS cell populations were stably transfected with a modified, TET regulated, CMV based expression vector (pCDNA4/TO, Invitrogen) encoding the ZFP fusion proteins; ZFP-A G9A cat domain or ZFP-A Suv39H1 Del76, respectively. Doxycycline-dependent

20   expression of either ZFP fusion protein induced repression of VEGF-A as assayed via ELISA and repression of IGF2 and H19 transcription, as assayed by Taqman (assays performed 48 hours post Dox induction).

As shown in Figures 5 through 10, various ZFP-CME fusions repress the expression of VEGF-A, IGF-2 and H19 reporter genes.

25

        Example 7: Assay for Compounds

1. Cells stably transfected with a ZFP-CME are plated at a specified cell density in a 96 well format.

2. Induce ZFP-fusion protein expression with addition of DOX to media. At the same time

30   a compound from the compound library is added to the plates at one or more concentrations and the plates are incubated for required time period.

3. Cells are lysed and RNA extracted, or protein based assays are used to determine the effects of the compounds screened upon the regulation of the endogenous reporter gene(s) via the ZFP-fusion protein, versus a DOX induced control. Compounds appearing to give a significant

35   modulation of ZFP fusion protein function selected for further testing.
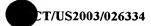
Example 8: Assay for modulators of oncogenic fusion proteins

cDNA is prepared from RNA extracted from leukemic cell lines which contain and express the chimeric translocation of interest (e.g. PML/RARα), using the thermoscript RT-PCR system (GibcoBRL) according to manufacturers recommendations. The DNA encoding the PML/RARα

5   fusion protein is generated via PCR from the cDNA pool using Platinum Taq High Fidelity DNA polymerase (Invitrogen) and sequence confirmed. Using PCR and standard molecular biology techniques, the DNA binding domain of RARα portion of this chimera is replaced with a ZFP DNA binding domain, which is itself targeted to one or more endogenous reporters. Using transient transfections of expression plasmids of this construct into human cell lines, the repressive

10   activity of this construct upon the transcriptional activity of the endogenous reporter gene(s)can be confirmed via RNA or protein based assays. These ZFP chimeras can then also be used to create stable, inducible cell lines for screening purposes. Such an approach, *i.e.*, replacement of the DNA binding domain of the leukemic translocation with a ZFP DNA binding domain, allows a system to be constructed to directly assay for inhibitors of the aberrant activity of these oncogenic chimeric

15   proteins, and is applicable to the development of inhibitors for a series of these chimeric proteins. In the presence of a compound which inhibits the repressive activity, increased expression of the reporter gene(s) is observed.

Example 9: Targeted antagonism of nuclear hormone receptor-activated gene
20   **expression by a ZFP-PAD V fusion protein**

Peptidylarginine deiminase V (PAD V) deiminates arginine residues, thereby converting them to citrulline, in histones H3 and H4. The particular arginine residues acted upon by PAD V are normally methylated (in nucleosomes located in or near the gene) by CARM1 and/or PRMT1 during the process of transcriptional activation of gene expression mediated by nuclear hormone

25   receptors (NHRs) such as, for example, ER alpha. To determine whether targeted PAD V activity antagonizes NHR-mediated gene activation, fusions of PAD V and a number of engineered zinc finger binding domains were constructed.
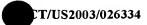
The zinc finger binding domains were targeted to the human vascular endothelial growth factor (VEGF) gene and are denoted VOP32E and VOP 30A. *See* co-owned US Patent application

30   US2003/0021776 for their amino acid sequences and DNA target sites. Methods for the design and synthesis of zinc finger binding domains and fusion proteins are disclosed, for example, in co-owned US Patent No. 6,453,242.

For these experiments, 293 cells were seeded into 12 well dishes at a density of 200,000 cells per well, 24 hours prior to transfection. Cells were then transfected with 1.5ug of the

35   plasmids indicated in Figure 11 using 8ul of Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. 6 hours post-transfection the medium was changed, and certain

cells (as indicated) were stimulated with Beta Estradiol (B Est) 24 hours post transfection. 72
hours post-transfection cells were harvested and total RNA extracted using the High pure RNA
isolation kit (Roche). Real-time quantitative PCR (Taqman) was then performed using this RNA
assaying the relative expression of the VEGF-A target gene compared with the expression of the
5   GAPDH control gene. Results are shown as the ratio of VEGF-A to GAPDH expression.

A fusion between the ER alpha ligand binding domain (amino acids 308-595) and a VEGF-
targeted zinc finger binding domain (VOP 32B) was constructed and shown to activate VEGF
transcription in a beta-estradiol-dependent fashion. *See* Figure 11, lane 4 compared to lanes 7
(empty vector control) and 9 (mock-transfected cells). When a PAD V fusion to a different VEGF-
10   targeted ZFP (VOP 30) was expressed at the same time as the ER alpha-ZFP fusion, estradiol-
stimulated activation of VEGF transcription was reduced. *See* Figure 11, lane 5. VEGF-targeted
ZFP-PAD V fusions, in the absence of ER alpha, had no effect on VEGF transcription (compare
lanes 1 and 2 with lanes 7-9). Moreover, a ZFP-PAD V fusion targeted to a sequence other than
the VEGF gene, had no effect on either basal VEGF transcription (lane 3) or ER alpha-stimulated
15   activation of VEGF transcription (lane 6). Accordingly, NHR-stimulated activation of gene
expression can be reduced or blocked by targeting the activity of a histone modifying enzyme such
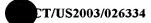as, for example, PAD V to the gene.

References

20   1.   Berger SL: **Histone modifications in transcriptional regulation.** *Curr Opin Genet Dev*
        2002; **12**:142-148.

    2.   Zhang Y, Reinberg D: **Transcription regulation by histone methylation: interplay**
        between different covalent modifications of the core histone tails. *Genes Dev* 2001;
        15:2343-2360.

25   3.   Kouzarides T: **Histone methylation in transcriptional control.** *Curr Opin Genet Dev*
        2002; **12**:198-209.

    4.   Strahl BD, Allis CD: **The language of covalent histone modifications.** *Nature* 2000;
        403:41-45.

    5.   Jenuwein T, Allis CD: **Translating the histone code.** *Science* 2001; **293**:1074-1080.

30   6.   Murray K: **The occurrance of N-methyl lysine in histones.** *Biochemistry* 1964; **3**:10-15.

    7.   Nishioka K, Reinberg D: **Transcription. Switching partners in a regulatory tango.**
        Science 2001; **294**:2497-2498.

    8.   Ma H, Baumann CT, Li H, Strahl BD, Rice R, Jelinek MA, Aswad DW, Allis CD, Hager
        GL, Stallcup MR: **Hormone-dependent, CARM1-directed, arginine-specific**
35   **methylation of histone H3 on a steroid-regulated promoter.** *Curr Biol* 2001; **11**:1981-
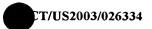   1985.

9.   Bauer UM, Daujat S, Nielsen SJ, Nightingale K, Kouzarides T: **Methylation at arginine 17 of histone H3 is linked to gene activation.** *EMBO Rep* 2002; 3:39-44.

10.   Wang H, Huang ZQ, Xia L, Feng Q, Erdjument-Bromage H, Strahl BD, Briggs SD, Allis CD, Wong J, Tempst P, Zhang Y: **Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor.** *Science* 2001; 293:853-857.

11.   Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P, Reinberg D: **Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation.** *Genes Dev* 2002; 16:479-489.

12.   Strahl BD, Ohba R, Cook RG, Allis CD: **Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena.** *Proc Natl Acad Sci U S A* 1999; 96:14967-14972.

13.   Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, Winston F, Allis CD: **Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae.** *Genes Dev* 2001; 15:3286-3295.

14.   Bryk M, Briggs SD, Strahl BD, Curcio MJ, Allis CD, Winston F: **Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in S. cerevisiae by a Sir2-independent mechanism.** *Curr Biol* 2002; 12:165-170.

15.   Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T: **Regulation of chromatin structure by site-specific histone H3 methyltransferases.** *Nature* 2000; 406:593-599.

16.   O'Carroll D, Scherthan H, Peters AH, Opravil S, Haynes AR, Laible G, Rea S, Schmid M, Lebersorger A, Jerratsch M, Sattler L, Mattei MG, Denny P, Brown SD, Schweizer D, Jenuwein T: **Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression.** *Mol Cell Biol* 2000; 20:9423-9433.

17.   Reuter G, Spierer P: **Position effect variegation and chromatin proteins.** *Bioessays* 1992; 14:605-612.

18.   Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G: **The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes.** *Embo J* 1994; 13:3822-3831.

19.   Ivanova AV, Bonaduce MJ, Ivanov SV, Klar AJ: **The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast.** *Nat Genet* 1998; 19:192-195.

20.   Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI: **Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.** *Science* 2001; 292:110-113.

21.   Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T: **Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.** *Nature* 2001; 410:120-124.

22.   Loyola A, LeRoy G, Wang YH, Reinberg D: **Reconstitution of recombinant chromatin establishes a requirement for histone-tail modifications during chromatin assembly and transcription.** *Genes Dev* 2001; 15:2837-2851.

23.   Nielsen SJ, Schneider R, Bauer UM, Bannister AJ, Morrison A, O'Carroll D, Firestein R, Cleary M, Jenuwein T, Herrera RE, Kouzarides T: **Rb targets histone H3 methylation and HP1 to promoters.** *Nature* 2001; 412:561-565.

24.   Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y: **A complex with chromatin modifiers that occupies E2F- and Myc- responsive genes in G0 cells.** *Science* 2002; 296:1132-1136.

25.   Tachibana M, Sugimoto K, Fukushima T, Shinkai Y: **Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3.** *J Biol Chem* 2001; 276:25309-25317.

26.   Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ, 3rd: **SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins.** *Genes Dev* 2002; 16:919-932.

27.   Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP, Case CC: **Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene.** *J Biol Chem* 2000; 275:33850-33860.

28.   Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, Zhong X, Lee YL, Eisenberg SP, Spratt SK, Case CC, Wolffe AP: **Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A.** *J Biol Chem* 2001; 276:11323-11334.

29.   Ren D, Collingwood TN, Rebar EJ, Wolffe AP, Camp HS: **PPARgamma knockdown by engineered transcription factors: exogenous PPARgamma2 but not PPARgamma1 reactivates adipogenesis.** *Genes Dev* 2002; 16:27-32.

30. Wang H, Cao R, Xia L, Erdjument-Bromage H, Borchers C, Tempst P, Zhang Y: **Purification and functional characterization of a histone H3-lysine 4- specific methyltransferase.** *Mol Cell* 2001; 8:1207-1217.

31. Melcher M, Schmid M, Aagaard L, Selenko P, Laible G, Jenuwein T: **Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression.** *Mol Cell Biol* 2000; 20:3728-3741.

32. Aagaard L, Laible G, Selenko P, Schmid M, Dorn R, Schotta G, Kuhfittig S, Wolf A, Lebersorger A, Singh PB, Reuter G, Jenuwein T: **Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3- 9 encode centromere-associated proteins which complex with the heterochromatin component M31.** *Embo J* 1999; 18:1923-1938.

33. Urnov FD, Wolffe AP: **A necessary good: nuclear hormone receptors and their chromatin templates.** *Mol Endocrinol* 2001; 15:1-16.

34. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T: **Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.** *Nature* 2001; 410:116-120.

35. Snowden AW, Anderson LA, Webster GA, Perkins ND: **A novel transcriptional repression domain mediates p21(WAF1/CIP1) induction of p300 transactivation.** *Mol Cell Biol* 2000; 20:2676-2686.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.